Detection boundary in sparse regression

Yuri I. Ingster, Alexandre Tsybakov, Nicolas Verzelen

Detection boundary in sparse regression

Ingster, Yu. I. [*], Tsybakov, A.B. [†], and Verzelen, N. [‡]

September 9, 2010

Abstract

We study the problem of detection of a p-dimensional sparse vector of parameters in the linear regression model with Gaussian noise. We establish the detection boundary, i.e., the necessary and sufficient conditions for the possibility of successful detection as both the sample size n and the dimension p tend to the infinity. Testing procedures that achieve this boundary are also exhibited. Our results encompass the high-dimensional setting ($p \gg n$). The main message is that, under some conditions, the detection boundary phenomenon that has been proved for the Gaussian sequence model, extends to high-dimensional linear regression. Finally, we establish the detection boundaries when the variance of the noise is unknown. Interestingly, the detection boundaries sometimes depend on the knowledge of the variance in a high-dimensional setting.

Mathematics Subject Classifications: Primary 62J05, Secondary 62G10, 62H20, 62G05, 62G08, 62C20, 62G20.
Key Words: High-dimensional regression, detection boundary, sparse vectors, sparsity, minimax hypothesis testing.

1 Introduction

We consider the linear regression model with random design:

$$Y_i = \sum_{j=1}^{p} \theta_j X_{ij} + \xi_i, \;\; i = 1, ..., n, \tag{1.1}$$

where $\theta_j \in \mathbb{R}$ are unknown coefficients, ξ_i are i.i.d. $\mathcal{N}(0, \sigma^2)$ random variables, X_{ij} are random variables, which are identically distributed, and $(X_{ij}, 1 \le i \le n)$ are independent for any fixed j with $EX_{ij} = 0$, $EX_{ij}^2 = 1$. We study separately the

[*]St.Petersburg State Electrotechnical University, 5, Prof. Popov str., 197376 St.Petersburg, Russia
[†]CREST, Timbre J340 3, av. Pierre Larousse, 92240 Malakoff cedex, France
[‡]INRA, UMR 729 MISTEA, 2 place Pierre Viala, F-34060 Montpellier, France

settings with known $\sigma > 0$ (then assuming that $\sigma = 1$ without loss of generality) and unknown $\sigma > 0$. We also assume that X_{ij}, $1 \le j \le p$, $1 \le i \le n$, are independent of $\xi_i, 1 \le i \le n$.

Based on the observations $Z = (X, Y)$ where $X = (X_{ij}, 1 \le j \le p,\ 1 \le i \le n)$, and $Y = (Y_i, 1 \le i \le n)$, we consider the problem of detecting whether the signal $\theta = (\theta_1, \dots, \theta_p)$ is zero (i.e., we observe the pure noise) or θ is some sparse signal, which is sufficiently well separated from 0. Specifically, we state this as a problem of testing the hypothesis $H_0 : \theta = 0$ against the alternative

$$H_{k,r} : \theta \in \Theta_k(r) = \{\theta \in \mathbb{R}^p_k : \|\theta\| \ge r\},$$

where $\mathbb{R}^p_k$ denotes the ℓ_0 ball in $\mathbb{R}^p$ of radius k, $\|\cdot\|$ is the Euclidean norm, and $r > 0$ is a separation constant.

The smaller is r, the harder is to detect the signal. The question that we study here is: What is the *detection boundary*, i.e., what is the smallest separation constant r such that successful detection is still possible? The problem is formalized in an asymptotic minimax sense, cf. Section 2 below. This question is closely related to the previous work by several authors on detection and classification boundaries for the Gaussian sequence model [4, 8, 9, 10, 11, 12, 13, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24]. These papers considered model (1.1) with $p = n$ and $X_{ij} = \delta_{ij}$, where δ_{ij} is the Kronecker delta, or replications of such a model (in classification setting). The main message of the present work is that, under some conditions, the detection boundary phenomenon similar to the one discussed in those papers, extends to linear regression. Our results cover the high-dimensional $p \gg n$ setting.

We now give a brief summary of our findings under the simplifying assumption that all the regressors X_{ij} are i.i.d. standard Gaussian. We consider the asymptotic setting where $p \to \infty$, $n \to \infty$ and $k = p^{1-\beta}$ for some $\beta \in (0, 1)$. The results are different for moderately sparse alternatives ($0 < \beta < 1/2$) and highly sparse alternatives ($1/2 < \beta < 1$). We show that for moderately sparse alternatives the detection boundary is of the order of magnitude

$$r \asymp \frac{p^{1/4}}{\sqrt{n}} \wedge \frac{1}{n^{1/4}}, \tag{1.2}$$

whereas for highly sparse alternatives ($1/2 < \beta < 1$) it is of the order

$$r \asymp \sqrt{\frac{k \log p}{n}} \wedge \frac{1}{n^{1/4}}. \tag{1.3}$$

This solves the problem of optimal rate in detection boundary for all the range of values (p, n). Furthermore, for highly sparse alternatives under the additional assumption

$$p^{1-\beta} \log(p) = o(\sqrt{n}) \tag{1.4}$$

we obtain the sharp detection boundary, i.e., not only the rate but also the exact constant. This sharp boundary has the form

$$r = \varphi(\beta) \sqrt{\frac{k \log p}{n}}, \tag{1.5}$$

where

$$\varphi(\beta) = \begin{cases} \sqrt{2\beta - 1}, & 1/2 < \beta \leq 3/4, \\ \sqrt{2}(1 - \sqrt{1-\beta}), & 3/4 < \beta < 1. \end{cases} \tag{1.6}$$

The function $\varphi(\cdot)$ here is the same as in the above mentioned detection and classification problems, as first introduced in [15]. We also provide optimal testing procedures. In particular, the sharp boundary (1.5)-(1.6) is attained on the Higher Criticism statistic.

One of the applications of this result is related to transmission of signals under compressed sensing, cf. [7, 5]. Assume that a sparse high-dimensional signal θ is coded using compressed sensing with i.i.d. Gaussian X_{ij} and then transmitted through a noisy channel. Observing the noisy outputs Y_i, we would like to detect whether the signal was indeed transmitted. For example, this is of interest if several signals appear in consecutive time slots but some slots contain no signal. Then the aim is to detect informative slots. Our detection boundary (1.5) specifies the minimal energy of the signal sufficient for detectability. We note that $\varphi(\cdot) < \sqrt{2}$, so that successful detection is possible for rather weak signals whose energy is under the threshold $\sqrt{2k\log(p)/n}$. This can be compared with the asymptotically optimal recovery of sparsity pattern (RSP) by the Lasso in the same Gaussian model as ours [29, 30]. Observe that the RSP property is stronger than detection (i.e., it implies correct detection) but [29] defines the alternative by $\{\theta \in \mathbb{R}^p_k : |\theta_j| \geq c\sqrt{\log(p)/n}, \forall j\}$ for some constant $c > 2$, which is better separated from the null than our alternative $\Theta_k(r)$. Thresholds that are larger in order of magnitude are required if one would like to perform detection based on estimation of the values of coefficients in the ℓ_2 norm [3, 5].

In many applications, the variance of the noise ξ is unknown. Does the problem of detection become more difficult in this case? In order to answer this question, we investigate the detection boundaries in the unknown variance setting. Related work [27, 28] develop minimax bounds for detection in model (1.1) under assumptions different from ours and under unknown variance. However, [27] does not provide a sharp boundary. Here, we prove that for $\beta \in (1/2, 1)$ and $k\log(p) \ll \sqrt{n}$, the detection boundaries are the same for known and unknown variance. In contrast, when $k\log(p) \gg \sqrt{n}$, the detection boundary is much larger in the case of unknown variance than for known variance. We also provide an optimal testing procedure for unknown variance.

After we have obtained our results, we became aware of the interesting parallel unpublished work of Arias-Castro et al. [2]. There the authors derive the detection boundary in model (1.1) with known variance of the noise for both fixed and random design. Their approach based on the analysis of the Higher Criticism shares some similarities with our work. When the variables X_{ij} are i.i.d. standard normal and the variance is known, we can directly compare our results with [2]. In [2] the detection boundaries analogous to (1.2) and (1.3) do not contain the minimum with the $n^{-1/4}$ term, because they are proved in a smaller range of values (p, n) where this term disappears. In particular, the conditions in [2] exclude the high-dimensional case $p \gg n$. We also note that, due to the constraints on the classes of matrices X, [2] obtains the sharp boundary (1.5)-(1.6) under the condition $p^{1-\beta}(\log(p))^2 = o(\sqrt{n})$ which is more restrictive than our condition (1.4). The other difference is that [2] does not treat the case of unknown variance of the noise.

Below we will use the following notation. We write $Z = (X, Y)$ where $X = (X_{ij}, 1 \le j \le p,\ 1 \le i \le n)$, and $Y = (Y_i, 1 \le i \le n)$ are the observations satisfying (1.1). Let P_θ be the probability measure that corresponds to observations Z, $P_{\theta,i}$ be those corresponding to observations $Z_i = (X_{i1}, ..., X_{ip}, Y_i)$ with fixed $i = 1, ..., n$, and $P_X, P_{X,i}$ be the probability measures corresponding to observations X or $X^{(i)} = (X_{i,j}, 1 \le j \le p)$. We denote by P_θ^X and $P_{\theta,i}^X$ the conditional distributions of Y given X and of Y_i given $X^{(i)}$ respectively. The corresponding expectations are denoted by E_θ^X and $E_{\theta,i}^X$. Clearly,

$$P_\theta(dZ) = P_\theta^X(dY)P_X(dX),\ P_{\theta,i}(dZ_i) = P_{\theta,i}^X(dY)P_{X,i}(dX^{(i)}), \tag{1.7}$$

and

$$P_\theta(dZ) = \prod_{i=1}^n P_{\theta,i}(dZ_i).$$

We denote by $X_j \in \mathbb{R}^n$ the jth column of matrix $X = (X_{ij})$, and set

$$(X_j, X_l) = \sum_{i=1}^n X_{ij}X_{il}, \quad \|X_j\|^2 = (X_j, X_j).$$

2 Detection problem

For $\theta \in \mathbb{R}^p$, we denote by $M(\theta) = \sum_{j=1}^p 1\!\!1_{\{\theta_j \ne 0\}}$ the number of non-zero components of θ, where $1\!\!1_{\{A\}}$ is the indicator function. As above, let $\mathbb{R}_k^p$, $1 \le k \le p$, denote the ℓ_0 ball in $\mathbb{R}^p$ of radius k, i.e., the subset of $\mathbb{R}^p$ that consists of vectors θ with $M(\theta) \le k$, or equivalently, $\theta \in \mathbb{R}_k^p$ contains no more than k nonzero coordinates. In particular $\mathbb{R}_p^p = \mathbb{R}^p$. Recall the notation $\Theta_k(r) = \{\theta \in \mathbb{R}_k^p : \|\theta\| \ge r\}$.

We consider the problem of testing the hypothesis $H_0 : \theta = 0$ against the alternative $H_{k,r} : \theta \in \Theta_k(r)$. In this paper we study the asymptotic setting where $p \to \infty$, $n \to \infty$ and $k = p^{1-\beta}$. The coefficient $\beta \in [0, 1]$ is called the sparsity index. We assume in this section that σ^2 is known. Modifications for the case of unknown variance are discussed in Section 4.2.1.

We call a test any measurable function $\psi(Z)$ with values in $[0, 1]$. For a test ψ, let $\alpha(\psi) = E_0(\psi)$ be the type I error probability and $\beta(\psi, \theta) = E_\theta(1 - \psi)$ be the type II error probability for the alternative $\theta \in \Theta \subset \mathbb{R}^p$. We set

$$\beta(\psi) = \beta(\psi, \Theta) = \sup_{\theta \in \Theta} \beta(\psi, \theta), \quad \gamma(\psi) = \gamma(\psi, \Theta) = \alpha(\psi) + \beta(\psi, \Theta)\ .$$

We denote by $\beta(\alpha) = \beta_{n,p,k}(\alpha, r)$ the minimax type II error probability for a given level $\alpha \in (0, 1)$,

$$\beta(\alpha) = \inf_{\psi:\alpha(\psi) \le \alpha} \beta(\psi, \Theta_k(r)), \quad 0 \le \beta(\alpha) \le 1 - \alpha\ .$$

Accordingly, we denote by $\gamma = \gamma_{n,p,k}(r)$ the minimax total error probability in the hypothesis testing problem:

$$\gamma = \inf_{\psi} \gamma(\psi, \Theta_k(r)),$$

where the infimum is taken over all tests ψ. Clearly,

$$\gamma = \inf_{\alpha \in (0,1)} (\alpha + \beta(\alpha)), \quad 0 \le \gamma \le 1.$$

The aim of this paper is to establish the asymptotic detection boundary, i.e., the conditions on the separation constant $r = r_{n,p,k}$, which delimit the zone where $\gamma_{n,p,k}(r) \to 1$ (indistinguishability) from that where $\gamma_{n,p,k}(r) \to 0$ (distinguishability). The distinguishability is equivalent to $\beta(\alpha) \to 0, \ \forall \ \alpha \in (0,1)$. We are interested in tests $\psi = \psi_{n,p}$ or $\psi_\alpha = \psi_{n,p,\alpha}$ such that either $\gamma(\psi) \to 0$ or $\alpha(\psi_\alpha) \le \alpha + o(1)$, and $\beta(\psi_\alpha) \to 0$. Here and later the limits are taken as $p \to \infty, \ n \to \infty$ unless otherwise stated.

3 Assumptions on X

We will use at different instances some of the following conditions on the random variables X_{ij}.

A1. *The random variables X_{ij} are uncorrelated, i.e., $EX_{ij}X_{il} = 0$ for all $1 \le j < l \le p$.*

A2. *The random variables X_{ij}, $1 \le j \le p, \ 1 \le i \le n$, are independent.*

A3. *The random variables X_{ij}, $1 \le j \le p, \ 1 \le i \le n$, are i.i.d. standard Gaussian: $X_{ij} \sim \mathcal{N}(0,1)$.*

Let $U_j, \ 1 \le j \le p$ be random variables such that we have equality in distribution $\mathcal{L}(U_j, U_l) = \mathcal{L}(X_{ij}, X_{il}), \quad 1 \le j < l \le p$. We will need the following technical assumptions.

B1.
$$\max_{1 \le j < l \le p} E((U_j U_l)^4) = O(1) \ . \tag{3.1}$$

B2. *There exists $h_0 > 0$ such that $\max_{1 \le j \le l \le p} E(\exp(hU_jU_l)) < \infty$ for $|h| < h_0$, and*

$$\log^3(p) = o(n). \tag{3.2}$$

B3. *There exists $m \in \mathbb{N}$ such that $\max_{1 \le j \le l \le p} E(|U_jU_l|^m) < \infty$, and*

$$\log^2(p) p^{4/m} = o(n). \tag{3.3}$$

Assumption **B1** implies that

$$\max_{1 \le j < l \le p} E(|U_jU_l|^m) \ = \ O(1), \quad m = 2, 3, 4 \ . \tag{3.4}$$

In particular, Assumption **B1** holds true under **A2** if

$$\max_{1\le j\le p} E(U_j^4) = O(1). \tag{3.5}$$

If $(X_{ij}X_{il}, i = 1, \dots, n)$ are independent zero-mean random variables, we have (cf. [25], p. 79):

$$E|(X_j, X_l)|^m \le C(m)n^{m/2-1}\sum_{i=1}^{n} E(|X_{ij}X_{il}|^m), \quad m > 2 \ .$$

This and (3.4) yield

$$\sum_{1\le j<l\le p} E(|(X_j, X_l)|^m) \ = \ O(n^{m/2}p^2), \quad m = 2, 3, 4 \ . \tag{3.6}$$

Finally, Assumptions **B1** and **B2** hold true under **A3** and (3.2).

4 Main results

4.1 Detection boundary under known variance

For this case we suppose $\sigma = 1$ without loss of generality.

4.1.1 Lower bounds

We first present the lower bounds on the detection error, i.e., the indistinguishability conditions. We assume that $k = p^{1-\beta}, \beta \in (0, 1)$. Indistinguishability conditions consist of two joint conditions on the radius $r = r_{np}$. The first one is

$$r_{np}^2 = o(n^{-1/2}). \tag{4.1}$$

The second condition differs according to whether $\beta \le 1/2$ or $\beta > 1/2$. If $\beta \le 1/2$ (i.e. $p = O(k^2)$), which corresponds to moderate sparsity, we require that

$$r_{np}^2 = o(\sqrt{p}/n) \ . \tag{4.2}$$

The case $\beta > 1/2$ (i.e. $k^2 = o(p)$) corresponds to high sparsity. We define x_{np} by $r_{np} = x_{n,p}\sqrt{k\log(p)/n}$. Then, we require that

$$\limsup(x_{n,p} - \varphi(\beta)) < 0, \tag{4.3}$$

where $\varphi(\beta)$ is defined in (1.6). Clearly, condition (4.3) implies $r_{np}^2 = O(k\log(p)/n)$, which is stronger than (4.2) when $\beta > 1/2$.

Theorem 4.1 *Assume* **A1**, **B1**, $k = p^{1-\beta}$ *and either* **B2** *or* **B3**. *We also require that* r_{np} *satisfies (4.1) and either (4.2) (for* $\beta \in (0, 1/2]$*) or (4.3) (for* $\beta \in (1/2, 1)$*). Then, asymptotic distinguishability is impossible, i.e.,* $\gamma_{n,p,k}(r_{np}) \to 1$.

Remark 4.1 *This theorem can be extended to non-random design matrix X. Inspection of the proof shows that, instead of* **B1**, *we only need the assumption: For some $B_{n,p}$ tending to ∞ slowly enough,*

$$\sum_{1\le j<l\le p} |(X_j, X_l)|^m < B_{n,p} n^{m/2} p^2, \quad m = 2, 3, 4. \tag{4.4}$$

Indeed, **B1** *is used in the proofs only to assure that (4.4) holds true with P_X-probability tending to 1 (this is deduced from assumption* **B1** *and (3.6)).*

Also instead of **B2** *and* **B3**, *we can assume that there exists $\eta_{n,p} \to 0$ such that*

$$r_{np}^2 \max_{1\le j<l\le p} |(X_j, X_l)| < \eta_{n,p} k, \quad \max_{1\le j\le p} |\|X_j\|^2 - n| < \eta_{n,p} n. \tag{4.5}$$

Under **B2**, **B3**, *relations (4.5) hold with P_X-probability tending to 1, see Corollary 7.1.*

The result of the theorem remains valid for non-random matrices X satisfying (4.4) and (4.5).

4.1.2 Upper bounds

In order to construct a test procedure that achieves the detection boundary, we combine several tests.

First, we study the widest non-sparse case $k = p$, i.e., we consider $\Theta_p(r) = \{\theta \in \mathbb{R}^p : \|\theta\| \ge r\}$. Consider the statistic

$$t_0 = (2n)^{-1/2} \sum_{i=1}^{n} (Y_i^2 - 1), \tag{4.6}$$

which is the H_0-centered and normalized version of the classical χ_n^2-statistic $\sum_{i=1}^n Y_i^2$. The corresponding tests ψ_α^0 and ψ^0 are of the form:

$$\psi_\alpha^0 = 1\!\!\mathrm{I}_{t_0 > u_\alpha}, \quad \psi^0 = 1\!\!\mathrm{I}_{t_0 > T_{np}}$$

where $\alpha \in (0, 1)$, u_α is the $(1-\alpha)$-quantile of the standard Gaussian distribution and T_{np} is any sequence satisfying $T_{np} \to \infty$.

Theorem 4.2 *For all $\alpha \in (0, 1)$, we have:*

(i) Type I errors satisfy $\alpha(\psi_\alpha^0) = \alpha + o(1)$ and $\alpha(\psi^0) = o(1)$.

(ii) Type II errors. Assume **A2** *and* **B1** *and consider a radius r_{np} such that $nr_{np}^4 \to \infty$. Then, we have $\beta(\psi_\alpha^0, \Theta_p(r_{np})) \to 0$. If T_{np} is chosen such that $\limsup T_{np} n^{-1/2} r_{np}^{-2} < 1$, then $\beta(\psi^0, \Theta_p(r_{np})) \to 0$.*

Recall that we can replace **B1** by (3.5) under **A2**. If $nr_{np}^4 \to \infty$, then one can take T_{np} such that $\gamma_{n,p}(\psi^0, \Theta_p(r_{np})) \to 0$ under **A2**, **B1**. This upper bound corresponds to the part (4.1) of the detection boundary.

We now introduce a test ψ^1_α that achieves the second boundary (4.2). Consider the following kernel

$$K(Z_i, Z_k) = p^{-1/2} Y_i Y_k \sum_{j=1}^{p} X_{ij} X_{kj}.$$

The U-statistic t_1 based on the kernel K is defined by

$$t_1 = N^{-1/2} \sum_{1 \le i < k \le n} K(Z_i, Z_k), \quad N = n(n-1)/2.$$

Note that the U-statistic t_1 can be viewed as the H_0-centered and normalized version of the statistic $\chi^2_p = n \sum_{j=1}^{p} \hat{\theta}_j^2$ based on the estimators $\hat{\theta}_j = n^{-1} \sum_{i=1}^{n} Y_i X_{ij}$:

$$\chi^2_p = 2n^{-1} \sum_{j=1}^{p} \sum_{1 \le i < k \le n} Y_i Y_k X_{ij} X_{kj} + n^{-1} \sum_{j=1}^{p} \sum_{i=1}^{n} Y_i^2 X_{ij}^2.$$

Indeed, up to a normalization, the first sum is the U-statistic t_1, and moving off the second sum corresponds to centering.

Given $\alpha \in (0,1)$, we consider the test $\psi^1_\alpha = 1\!\!1_{t_1 > u_\alpha}$.

Theorem 4.3 *Assume* **A2** *and* **B1**. *For all* $\alpha \in (0,1)$ *we have:*

(i) Type I error satisfies: $\alpha(\psi^1_\alpha) = \alpha + o(1)$.

(ii) Type II errors. Assume that $p = o(n^2)$ *and consider a radius* r_{np} *such that* $n r^2_{np}/\sqrt{p} \to \infty$. *Then,* $\beta(\psi^1_\alpha, \Theta_p(r_{np})) \to 0$.

Remark 4.2 *Combining the tests* $\psi^0_{\alpha/2}$ *and* $\psi^1_{\alpha/2}$ *we obtain the test* $\psi^*_\alpha = \max(\psi^0_{\alpha/2}, \psi^1_{\alpha/2})$ *of asymptotic level not more than* α. *Moreover, it achieves* $\beta(\psi^1_\alpha, \Theta_p(r_{np})) \to 0$ *for any radius* r_{np} *satisfying*

$$r^2_{np} \gg \frac{\sqrt{p}}{n} \wedge \frac{1}{\sqrt{n}} \ .$$

We can omit the condition $p = o(n^2)$ *since the test* $\psi^0_{\alpha/2}$ *achieves the optimal rate for* $p \ge n$. *Combining this bound with Theorem 4.1, we conclude that* ψ^*_α *simultaneously achieves the optimal detection rate for all* $\beta \in (0, 1/2]$.

We now turn to testing in the highly-sparse case $\beta \in (1/2, 1)$. Here we use a version of "Higher Criticism Tests" (HC-tests, cf. [8]). Set

$$y_i = (Y, X_i)/\|Y\|, \quad 1 \le i \le p.$$

Let $q_i = P(|\mathcal{N}(0,1)| > |y_i|)$ be the p-value of the i-th component and let $q_{(i)}$ denote these quantities sorted in *increasing order*. We define the HC-statistic by

$$t_{HC} = \max_{i,\ q_{(i)} \le 1/2} \frac{\sqrt{p}(i/p - q_{(i)})}{\sqrt{q_{(i)}(1 - q_{(i)})}} \ . \tag{4.7}$$

Given a constant $a > 0$, the HC-test ψ^{HC} rejects H_0 when the statistic t_{HC} is larger than $(1+a)\sqrt{2 \log \log p}$.

Remark 4.3 *The cutoff* $1/2$ *in the definition (4.7) of* t_{HC} *can be replaced by any* $c \in (0,1)$.

Theorem 4.4 *Assume* **A3** *(*X_{ij} *are i.i.d. standard Gaussian).*
(i) Type I error satisfies $\alpha(\psi^{HC}) = o(1)$.
(ii) Type II error. Consider $k = p^{1-\beta}$ *with* $\beta \in (1/2,1)$ *and assume that* $k\log(p) = o(n)$. *Take a radius* $r_{np} = x_{np}\sqrt{k\log(p)/n}$ *such that* $\liminf(x_{np} - \varphi(\beta)) > 0$. *Then, we have* $\beta(\psi^{HC}, \Theta_k(r_{np})) \to 0$.

Remark 4.4 *If* $k\log(p) = o(n)$, *the HC-test asymptotically detects any* k*-sparse signal whose rescaled intensity* $r_{np}\sqrt{n/(k\log(p))}$ *is above the detection boundary* $\varphi(\beta)$.

Remark 4.5 *Assume* **A3**. *Combining the tests* ψ^0_α *and* ψ^{HC}, *we obtain the test* $\psi^{*,HC}_\alpha = \max(\psi^0_\alpha, \psi^{HC})$ *of asymptotic level not more than* α. *Moreover, it achieves* $\beta(\psi^1_\alpha, \Theta_k(r_{np})) \to 0$ *for any radius* $r_{n,p}$ *satisfying*

$$\liminf x_{n,p} \geq \varphi(\beta) \quad \text{or} \quad r^2_{np} \gg \frac{1}{\sqrt{n}} .$$

We can omit the condition $k\log(p) = o(n)$ *since the test* ψ^0_α *achieves the optimal rate for* $k\log(p) \gg \sqrt{n}$. *Combining this bound with Theorem 4.1, we conclude that* $\psi^{*,HC}_\alpha$ *simultaneously achieves the optimal detection rate for all* $\beta \in (1/2,1)$.

In conclusion, under Assumption **A3**, the test $\max(\psi^0_{\alpha/2}, \psi^1_{\alpha/2}, \psi^{HC})$ simultaneously achieves the optimal detection rate for all $\beta \in (0,1)$. The detection boundary is of the order of magnitude

$$r \asymp \sqrt{\frac{k\log p}{n}} \wedge \frac{1}{n^{1/4}} . \tag{4.8}$$

Furthermore, we establish the sharp detection boundary (i.e., with exact asymptotic constant) of the form

$$r = \varphi(\beta)\sqrt{\frac{k\log p}{n}}$$

for $\beta > 1/2$ and $k\log(p) = p^{1-\beta}\log(p) = o(\sqrt{n})$.

4.2 Detection boundary under unknown variance

4.2.1 Detection problem

Since the variance of the noise is now assumed to be unknown, the tests ψ under study should not require the knowledge of σ^2. The type I error probability is now taken uniformly over $\sigma > 0$:

$$\alpha^{un}(\psi) = \sup_{\sigma>0} E_{0,\sigma}(\psi) .$$

The type II error probability over an alternative $\Theta \subset \mathbb{R}^p$ is

$$\beta^{un}(\psi, \Theta) = \sup_{\theta \in \Theta, \sigma > 0} \beta(\psi, \theta\sigma, \sigma) = \sup_{\theta \in \Theta, \sigma > 0} E_{\theta\sigma,\sigma}(1 - \psi) \ . \tag{4.9}$$

Similarly to the setting with known variance, we consider the sum of the two errors:

$$\gamma^{un}(\psi, \Theta) = \alpha^{un}(\psi) + \beta^{un}(\psi, \theta).$$

Finally, the minimax total error probability in the hypothesis testing problem with unknown variance is

$$\gamma^{un}_{n,p,k}(r) = \inf_{\psi} \gamma^{un}(\psi, \Theta_k(r))$$

4.2.2 Lower bounds

Take $r_{np} = x_{n,p}\sqrt{k \log(p)/n}$. As in the case of known variance, we consider the condition

$$\limsup (x_{n,p} - \varphi(\beta)) < 0 \ . \tag{4.10}$$

Theorem 4.5 *Fix some $\beta > 1/2$ and assume* **A3**. *If Condition (4.10) holds and if $k \log(p) = o(n)$, then distinguishability is impossible, i.e., $\gamma^{un}_{n,p,k}(r_{np}) \to 1$.*

If $k \log(p)/n \to \infty$, then for any radius $r > 0$, distinguishability is impossible, i.e. $\gamma^{un}_{n,p,k}(r) \to 1$.

The detection boundary stated in Theorem 4.5 does not depend on the unknown σ^2. This is due to the definition (4.9) of the type II error probability $\beta^{un}(\psi, \Theta_k(r))$ that considers alternatives of the form $\sigma\theta$ with $\theta \in \Theta_k(r)$.

4.2.3 Upper bounds

The HC-test ψ^{HC} defined in (4.7) still achieves the optimal detection rate when the variance is unknown as shown by the next proposition.

Proposition 4.6 *Assume* **A3** *(X_{ij} are i.i.d. standard Gaussian).*

(i) Type I error satisfies $\alpha^{un}(\psi^{HC}) = o(1)$.

(ii) Type II error. Consider $k = p^{1-\beta}$ with $\beta \in (1/2, 1)$ and assume that $k \log(p) = o(n)$. Take a radius $r_{np} = x_{np}\sqrt{k \log(p)/n}$ such that $\liminf(x_{np} - \varphi(\beta)) > 0$. Then, we have $\beta^{un}(\psi^{HC}, \Theta_k(r_{np})) \to 0$.

In conclusion, in the setting with unknown variance we prove that the sharp detection boundary (i.e., with exact asymptotic constant) of the form

$$\varphi(\beta)\sqrt{\frac{k \log p}{n}}$$

holds for $\beta > 1/2$ and $k \log(p) = p^{1-\beta} \log(p) = o(n)$, i.e., for a larger zone of values (p, n) than for the case of known variance. However, this extension corresponds to (p, n) for which the rate itself is strictly slower than under the known variance. Indeed, if the variance σ^2 is known, as shown in Section 4.1, the detection boundary is of the order (4.8). Thus, there is an asymptotic difference in the order of magnitude of the two detection boundaries for $k \log(p) \gg \sqrt{n}$.

5 Proofs of the lower bounds

5.1 The prior

Take $c \in (0,1)$, $h = ck/p$, $b = r_{np}/c\sqrt{k}$, $a = b\sqrt{n}$. Note that the condition $r_{np}^2 = o(1/\sqrt{n})$ is equivalent to $b^4k^2n = o(1)$.

Let us consider a random vector $\theta = (\theta_j)$ with coordinates

$$\theta_j = b\varepsilon_j, \quad \text{where} \quad \varepsilon_j \in (0, +1, -1) \quad \text{iid},$$

such that

$$Prob(\varepsilon_j = 0) = 1 - h, \quad Prob(\varepsilon_j = +1) = Prob(\varepsilon_j = -1) = h/2.$$

This introduces a prior probability measure π_j on θ_j and the product prior measure $\pi = \prod_{j=1}^p \pi_j$ on θ. The corresponding expectation and variance operators will be denoted by E_π and Var_π.

Lemma 5.1 *Let $k \to \infty$. Then*

$$\pi(\theta \in \mathbb{R}_k^p,\ \|\theta\| \geq r) \to 1.$$

Proof. Observe that

$$\|\theta\|^2 = b^2 \sum_{j=1}^p \varepsilon_j^2,\ m(\theta) = \sum_{j=1}^p |\varepsilon_j|.$$

We have

$$E_\pi(\|\theta\|^2) = b^2ph = r^2/c, \quad E_\pi(m(\theta)) = ph = ck,$$

and

$$\mathrm{Var}_\pi(\|\theta\|^2) \leq phb^4 = r^4/(kc^3), \quad \mathrm{Var}_\pi(m(\theta)) \leq ph = ck.$$

Applying the Chebyshev inequality, we get with $C = c^{-1} > 1$,

$$\pi(\|\theta\|^2 < r^2) = \pi(E_\pi(\|\theta\|^2) - \|\theta\|^2 > r^2(C-1)) \leq \frac{\mathrm{Var}_\pi(\|\theta\|^2)}{r^4(C-1)^2} \to 0,$$

and similarly, $\pi(m(\theta) > k) \to 0$. $\square$

Lemma 5.1 implies that, in order to obtain asymptotic lower bounds for the minimax problem, we only have to study the Bayesian problem which corresponds to the prior π, see for instance [18], Proposition 2.9. Consider the mixture

$$P_\pi(dZ) = E_\pi P_\theta(dZ) = \int_{\mathbb{R}^p} P_\theta(dZ)\pi(d\theta)$$

and the likelihood ratio

$$L_\pi(Z) = \frac{dP_\pi}{dP_0}(Z).$$

In order to prove the lower bounds we only need to check that

$$L_\pi(Z) \to 1 \quad \text{in } P_0 - \text{probability}. \tag{5.1}$$

Consider $x = \limsup x_{n,p}$. If $\beta \le 1/2$, then $x = 0$ since $nb^2 = O(1)$. For $\beta > 1/2$, we take $c \in (0,1)$ such that $x_c = x/c < \varphi(\beta)$, which is possible as $x < \varphi(\beta)$. We will use the short notation x and a for x_c and $a_c = b\sqrt{n} = a/c$. We set

$$a_j = b\|X_j\|, \quad y_j' = (X_j, Y)/\|X_j\|, \quad x_j = a_j/\sqrt{\log(p)}, \quad T_j = a_j/2 + \log(h^{-1})/a_j,$$

which corresponds to $he^{-a_j^2 + a_j T} = 1$.

5.2 Study of the likelihood ratio L_π

First observe that by (1.7)

$$P_\pi(dZ) = P_X(dX)E_\pi\left(P_\theta^X(dY)\right), \quad L_\pi(Z) = E_\pi\left(\frac{dP_\theta^X}{dP_0^X}(Y)\right).$$

Note that conditional measure P_θ^X corresponds to observation of the Gaussian vector $\mathcal{N}(v, I_n)$ where $v = \sum_{j=1}^p \theta_j X_j$, I_n is the $n \times n$ identity matrix, and the likelihood ratio under the expectation is

$$\frac{dP_\theta^X}{dP_0^X}(Y) = \exp(-\|v\|^2/2 + (v, Y)) = g_\theta(Z)e^{-\Delta(X,\theta)},$$

where

$$g_\theta(Z) = \prod_{j=1}^p \exp(-\theta_j^2\|X_j\|^2/2 + \theta_j(X_j, Y)), \quad \Delta(X,\theta) = 2\sum_{1\le j<l\le p} \theta_j\theta_l(X_j, X_l). \tag{5.2}$$

Put

$$\Lambda(Z) = E_\pi\left(g_\theta(Z)\right) = \prod_{j=1}^p \left(1 - h + he^{-b^2\|X_j\|^2/2}\cosh(b(X_j, Y))\right).$$

We define $\eta_j = e^{-b^2\|X_j\|^2/2}\cosh(b(X_j, Y)) - 1$. Take now $\delta > 0$ and introduce the set

$$\Sigma_X = \{\theta \in \mathbb{R}^p : |\Delta(X,\theta)| \le \delta\}.$$

We can write

$$L_\pi(Z) = \int_{\mathbb{R}^p} g_\theta(Z)e^{-\Delta(X,\theta)}\pi(d\theta) \ge e^{-\delta}\int_{\Sigma_X} g_\theta(Z)\pi(d\theta) = e^{-\delta}\Lambda(Z)\pi_Z(\Sigma_X),$$

where $\pi_Z = \prod_{j=1}^p \pi_{Z,j}$ is the random probability measure on $\mathbb{R}^p$ with the density

$$\frac{d\pi_Z}{d\pi}(\theta) = \frac{g_\theta(Z)}{\Lambda(Z)} = \prod_{j=1}^p \frac{d\pi_{Z,j}}{d\pi_j}(\theta); \quad \frac{d\pi_{Z,j}}{d\pi_j}(\theta) = \frac{e^{-\theta_j^2\|X_j\|^2/2 + \theta_j(X_j,Y)}}{1 + h\eta_j}, \quad \theta_j \in \{0, \pm b\},$$

i.e., the measure $\pi_{Z,j}$ is supported at the points $\{0, b, -b\}$ and

$$\pi_{Z,j}(0) = \frac{1-h}{1+h\eta_j}, \ \pi_{Z,j}(\pm b) = \frac{h^{\pm}_{Z,j}}{2}, \quad h^{\pm}_{Z,j} = \frac{he^{d_j{}^{\pm}}}{1+h\eta_j},$$

where we set

$$d_j^{\pm} = -a_j^2/2 \pm a_j y_j' \ , \quad \eta_j = \frac{e^{d_j^+}}{2} + \frac{e^{d_j^-}}{2} - 1.$$

Proposition 5.1 *In P_0-probability,*

$$\pi_Z(\Sigma_X) \to 1. \tag{5.3}$$

Proof of Proposition 5.1 is given in Section 5.3.

Proposition 5.2 *In P_0-probability,*

$$\Lambda(Z) \to 1. \tag{5.4}$$

Proof of Proposition 5.2 is given in Section 5.4.

Propositions 5.1 and 5.2 imply that, for any $\delta > 0$,

$$P_0(Z : L_\pi(Z) > 1-\delta) \to 1.$$

Since $E_0 L_\pi = 1$ and $L_\pi(Z) \geq 0$, this yields $L_\pi \to 1$ in P_0-probability. This yields indistinguishability in the problem. $\square$

5.3 Proof of Proposition 5.1

5.3.1 Replacing the measure π_Z by $\tilde{\pi}_Z$

Let us consider the random measure $\tilde{\pi}_Z = \prod_{j=1}^p \tilde{\pi}_{Z,j}$, where $\tilde{\pi}_{Z,j}$ is supported at the points $\{0, b, -b\}$ and

$$\tilde{\pi}_{Z,j}(0) = 1 - \frac{q^+_{Z,j}}{2} - \frac{q^-_{Z,j}}{2}, \ \tilde{\pi}_{Z,j}(\pm b) = \frac{q^{\pm}_{Z,j}}{2},$$

where

$$q^{\pm}_{Z,j} = (h/2)e^{d_j{}^{\pm}} 1\!\!\mathrm{I}_{\mathcal{A}_j^{\pm}}, \quad \mathcal{A}_j^{\pm} = \{he^{d_j{}^{\pm}} < 1\} = \{\pm y_j' < T_j\}$$

and observe that the event $\mathcal{A}_j^{\pm}$ implies $q^{\pm}_{Z,j} \leq 1/2$, i.e, the measures $\tilde{\pi}_{Z,j}$ are correctly defined. We define the event $\mathcal{A} = \mathcal{A}_{np} = \cap_{j=1}^p (\mathcal{A}_j^+ \cap \mathcal{A}_j^-)$.

Lemma 5.2

$$P_0(\mathcal{A}_{n,p}) \to 1.$$

Proof. Denote A^c the complement of the event A. Since $y_j' \sim \mathcal{N}(0,1)$ under P_0, we have

$$P_0^X((\mathcal{A}_{n,p})^c) \leq \sum_{j=1}^p P_0^X((\mathcal{A}_j^+)^c) + P_0^X((\mathcal{A}_j^-)^c) = 2\sum_{j=1}^p \Phi(-T_j).$$

By Corollary 7.1 we get $a_j = b\|X_j\| \sim b\sqrt{n}$ uniformly in $1 \leq j \leq p$ in P_X-probability. By (7.1) this implies $\sum_{j=1}^p \Phi(-T_j) = o(1)$ in P_X-probability. $\square$

We can replace the measure π_Z by $\tilde{\pi}_Z$ in (5.3). This follows from the following lemma

Lemma 5.3 *In P_0-probability,*

$$E_{\tilde{\pi}_Z}|d\pi_Z/d\tilde{\pi}_Z - 1| \to 0. \tag{5.5}$$

Proof. Applying the equality $E_{\tilde{\pi}_Z}(d\pi_Z/d\tilde{\pi}_Z) = 1$ and the inequality $1 + x \leq e^x$, we get

$$\begin{aligned}
(E_{\tilde{\pi}_Z}|d\pi_Z/d\tilde{\pi}_Z - 1|)^2 &\leq E_{\tilde{\pi}_Z}(d\pi_Z/d\tilde{\pi}_Z - 1)^2 = E_{\tilde{\pi}_Z}(d\pi_Z/d\tilde{\pi}_Z)^2 - 1 \\
&= \prod_{j=1}^p E_{\tilde{\pi}_{Z,j}}\left(d\pi_{Z,j}/d\tilde{\pi}_{Z,j}\right)^2 - 1 \\
&= \prod_{j=1}^p (1 + E_{\tilde{\pi}_{Z,j}}(d\pi_{Z,j}/d\tilde{\pi}_{Z,j} - 1)^2) - 1 \\
&\leq \exp\left(\sum_{j=1}^p E_{\tilde{\pi}_{Z,j}}(d\pi_{Z,j}/d\tilde{\pi}_{Z,j} - 1)^2\right) - 1.
\end{aligned}$$

Consequently, we only have to prove that in P_0-probability,

$$H(Z) = \sum_{j=1}^p E_{\tilde{\pi}_{Z,j}}(d\pi_{Z,j}/d\tilde{\pi}_{Z,j} - 1)^2 \to 0.$$

Since $H(Z) \geq 0$, the last relation follows from

$$E_0^X(H) \to 0, \quad \text{in } P_X\text{-probability}$$

by Markov inequality. Observe that

$$E_{\tilde{\pi}_{Z,j}}(d\pi_{Z,j}/d\tilde{\pi}_{Z,j} - 1)^2 = \frac{(h_{Z,j}^+ - q_{Z,j}^+)^2}{2q_{Z,j}^+} + \frac{(h_{Z,j}^- - q_{Z,j}^-)^2}{2q_{Z,j}^-} + \frac{(h_{Z,j}^+ + h_{Z,j}^- - q_{Z,j}^+ - q_{Z,j}^-)^2}{2(2 - q_{Z,j}^+ - q_{Z,j}^-)}.$$

By Lemma 5.2, it is sufficient to study these terms under the event $\mathcal{A}$ which corresponds to $\max_{1\leq j\leq p} q_{Z,j}^{\pm} \leq 1/2$. Under this event, we have $h_{Z,j}^{\pm} = q_{Z,j}^{\pm}/\lambda_j$, $\lambda_j = 1 + q_{Z,j}^+ + q_{Z,j}^- - h$, and direct calculation gives

$$\frac{(h_{Z,j}^+ - q_{Z,j}^+)^2}{2q_{Z,j}^+} + \frac{(h_{Z,j}^- - q_{Z,j}^-)^2}{2q_{Z,j}^-} + \frac{(h_{Z,j}^+ + h_{Z,j}^- - q_{Z,j}^+ - q_{Z,j}^-)^2}{2(2 - q_{Z,j}^+ - q_{Z,j}^-)} = \frac{(q_{Z,j}^+ + q_{Z,j}^-)\Delta_j^2}{\lambda_j^2(2 - q_{Z,j}^+ - q_{Z,j}^-)},$$

where

$$\Delta_j = q^+_{Z,j} + q^-_{Z,j} - h = h(e^{d^+_j}\mathbb{1}_{\mathcal{A}^+_j} + e^{d^-_j}\mathbb{1}_{\mathcal{A}^-_j} - 2)/2.$$

Since $\max_{1\le j\le p} q^{\pm}_{Z,j} \le 1/2$, we only have to control the sum $\sum_{j=1}^p \Delta_j^2$.

$$E_0^X \mathbb{1}_{\mathcal{A}}\Delta_j^2 \;\le\; \frac{h^2}{2}\Big(e^{a_j^2}\Phi(T_j - 2a_j) + e^{-a_j^2} - 4\Phi(T_j - a_j) + 2\Big)\ .$$

CASE 1: $nb^2 = O(1)$. By corollary 7.1, $(a_j/(\sqrt{n}b) - 1) = o_{P_X}(1)$. Consequently, $\Phi(T_j - 2a_j) = 1 - o_{P_X}(p^{-2})$.

$$E_0^X\left[\mathbb{1}_{\mathcal{A}}\sum_{j=1}^p \Delta_j^2\right] \le \frac{ph^2}{2}\sinh^2(nb^2(1+o_{P_X}(1))/2) + o_{P_X}(1) = \frac{ph^2nb^2}{2} + o_{P_X}(1) = o_{P_X}(1)\ ,$$

since $r^2_{n,p} = o(\sqrt{p}/n)$.

CASE 2: $\limsup nb^2 = \infty$. This implies that $k^2 = o(p)$ and therefore $ph^2 = o(1)$.

$$E_0^X\left[\mathbb{1}_{\mathcal{A}}\sum_{j=1}^p \Delta_j^2\right] = \frac{ph^2}{2}e^{nb^2(1+o_{P_X}(1))} + o(1) = p^{-(2\beta-1)+x^2+o_{P_X}(1)} + o(1)\ .$$

Since $x < \varphi(\beta) \le \sqrt{2\beta - 1}$ for $\beta > 1/2$, this allows to conclude. □

5.3.2 Study of $E_{\tilde{\pi}_Z}\Delta^2$

By Lemma 5.2, the relation (5.3) follows from $\tilde{\pi}(\Sigma) \to 1$, in P_0-probability. Thus, we only need to check that in P_0-probability, $E_{\tilde{\pi}_Z}\Delta^2 \to 0$ for $\Delta = \Delta(X,\theta)$ defined by (5.2). By Markov inequality, the last relation follows from

$$E_0^X(E_{\tilde{\pi}_Z}\Delta^2))) \to 0, \quad \text{in } P_X\text{-probability.}$$

Let us introduce the events $\mathcal{X}_{n,p}$. Taking a positive family $\eta = \eta_{n,p} \to 0$, we set

$$\mathcal{X}^j = \{(\|X_j\|^2 - n) < \eta n\}, \quad \mathcal{X}^{ij} = \{\log(p)|(X_j, X_l)| < \eta n\}, \quad \mathcal{X}_{n,p} = \bigcap_{1\le j<l\le p}\left(\mathcal{X}^j \cap \mathcal{X}^{ij}\right).$$

It follows from Corollary 7.1 that, under assumptions **B2** or **B3** we can take $\eta = \eta_{n,p} \to 0$ such that $P_X(\mathcal{X}_{n,p}) \to 1$. We have

$$E_{\tilde{\pi}_Z}\Delta^2 = b^4 E_{\tilde{\pi}_Z}\left(\sum_{j_1,j_2,j_3,j_4=1}^p \theta_{j_1}\theta_{j_2}\theta_{j_3}\theta_{j_4}(X_{j_1}, X_{j_2})(X_{j_3}, X_{j_4})\right) = 2A_2 + 6A_3 + 24A_4,$$

where

$$A_2 = b^4 \sum_{1\le j_1<j_2\le p} E_{\tilde{\pi}_Z}\left(\epsilon^2_{j_1}\epsilon^2_{j_2}\right)(X_{j_1}, X_{j_2})^2, \tag{5.6}$$

$$A_3 = b^4 \sum_{1\le j_1<j_2<j_3\le p} E_{\tilde{\pi}_Z}\left(\epsilon^2_{j_1}\epsilon_{j_2}\epsilon_{j_3}\right)(X_{j_1}, X_{j_2})(X_{j_1}, X_{j_3}), \tag{5.7}$$

$$A_4 = b^4 \sum_{1\le j_1<j_2<j_3<j_4\le p} E_{\tilde{\pi}_Z}\left(\epsilon_{j_1}\epsilon_{j_2}\epsilon_{j_3}\epsilon_{j_4}\right)(X_{j_1}, X_{j_2})(X_{j_3}, X_{j_4}). \tag{5.8}$$

5.3.3 Expectation over $\tilde{\pi}_Z$ and over E_0^X

Let us define the variables η_k in $\{1,-1\}$. The expectations over $\tilde{\pi}_Z$ are of the form

$$E_{\tilde{\pi}_Z}\left(\epsilon_{j_1}^2\epsilon_{j_2}^2\right) = \frac{(q_{j_1}^+ + q_{j_1}^-)(q_{j_2}^+ + q_{j_2}^-)}{4} = \frac{1}{4}\sum_{\eta_1,\eta_2}\prod_{k=1}^{2} q_{j_k}^{\eta_k}, \tag{5.9}$$

$$E_{\tilde{\pi}_Z}\left(\varepsilon_{j_1}^2\varepsilon_{j_2}\varepsilon_{j_3}\right) = \frac{(q_{j_1}^+ + q_{j_1}^-)(q_{j_2}^+ - q_{j_2}^-)(q_{j_3}^+ - q_{j_3}^-)}{8} = \sum_{\eta_1,\eta_2,\eta_3}\frac{\eta_2\eta_3}{8}\prod_{k=1}^{3} q_{j_k}^{\eta_k} \tag{5.10}$$

$$E_{\tilde{\pi}_Z}\left(\varepsilon_{j_1}\varepsilon_{j_2}\varepsilon_{j_3}\varepsilon_{j_4}\right) = \frac{1}{16}\prod_{k=1}^{4}(q_{j_k}^+ - q_{j_k}^-) = \frac{1}{16}\sum_{\eta_1,\eta_2,\eta_3,\eta_4}\eta_1\eta_2\eta_3\eta_4\prod_{k=1}^{4} q_{j_k}^{\eta_k}. \tag{5.11}$$

Let us take the expectation E_0^X over Y of each of these expressions. We define the vector $V = b\sum_{k=1}^m \eta_k X_{j_k}$. Here, E_V^X refers to the expectation of Y over the Gaussian measure $\mathcal{N}(V, I_n)$. We derive that

$$\begin{aligned}
E_0^X\left(\prod_{k=1}^m q_{j_k}^{\eta_k}\right) &= \frac{h^m}{2^m}E_0^X\left(e^{-\frac{1}{2}\sum_{k=1}^m b^2\|X_{j_k}\|^2 + b(Y,\sum_{k=1}^m \eta_k X_{j_k})}\prod_{k=1}^m \mathbb{1}_{\{(Y,\eta_k X_{j_k})<T_{j_k}\|X_{j_k}\|\}}\right)\\
&= \frac{h^m}{2^m}\exp\left(b^2\sum_{1\le r<s\le m}\eta_r\eta_s(X_{j_r},X_{j_s})\right)E_0^X\left(e^{-\frac{1}{2}\|V\|^2+(Y,V)}\prod_{k=1}^m \mathbb{1}_{\{(Y,\eta_k X_{j_k})<T_{j_k}\|X_{j_k}\|\}}\right)\\
&= \frac{h^m}{2^m}\exp\left(b^2\sum_{1\le r<s\le m}\eta_r\eta_s(X_{j_r},X_{j_s})\right)E_V^X\left(\prod_{k=1}^m \mathbb{1}_{\{(Y,\eta_k X_{j_k})<T_{j_k}\|X_{j_k}\|\}}\right)\\
&= \frac{h^m}{2^m}\exp\left(b^2\sum_{1\le r<s\le m}\eta_r\eta_s(X_{j_r},X_{j_s})\right)P_{j_1,\ldots,j_m}(\eta),
\end{aligned}$$

where

$$\begin{aligned}
P_{j_1,\ldots,j_m}(\eta) &= E_V^X\left(\prod_{k=1}^m \mathbb{1}_{\{(Y,\eta_k X_{j_k})<T_{j_k}\|X_{j_k}\|\}}\right)\\
&= E_0^X\left(\prod_{k=1}^m \mathbb{1}_{\{(Y+V,\eta_k X_{j_k})<T_{j_k}\|X_{j_k}\|\}}\right)\\
&= E_0^X\left(\prod_{k=1}^m \mathbb{1}_{\{(Y,\eta_k X_{j_k})<T_{j_k}\|X_{j_k}\|-(V,\eta_k X_{j_k})\}}\right)\\
&= E_0^X\left(\prod_{k=1}^m \mathbb{1}_{\{\eta_k y'_{j_k}<T_{j_k}-(V,\eta_k X_{j_k})/\|X_{j_k}\|\}}\right).
\end{aligned}$$

Let us define

$$m_{j_k}(\eta) = \eta_k\sum_{s=1,\, s\neq k}^{m}\eta_s(X_{j_s},X_{j_k})/\|X_{j_k}\|, \quad z_k = \eta_k y'_{j_k}.$$

Then, $P_{j_1,\dots,j_m}(\eta)$ writes as

$$P_{j_1,\dots,j_m}(\eta) = P_0^X\left(z_1 < T_{j_1} - a_{j_1} - bm_{j_1}(\eta)\dots, z_m < T_{j_m} - a_{j_m} - bm_{j_m}(\eta)\right).$$

We have

$$E_0^X\left(\prod_{k=1}^{2} q_{j_k}^{\varepsilon_k}\right) = \frac{h^2}{4}\exp\left(\eta_1\eta_2 b^2(X_{j_1}, X_{j_2})\right) P_{j_1,j_2}(\eta), \tag{5.12}$$

$$E_0^X\left(\prod_{k=1}^{3} q_{j_k}^{\varepsilon_k}\right) = \frac{h^3}{8}\exp\left(b^2\sum_{1\le s<r\le 3}\eta_s\eta_r(X_{j_s}, X_{j_r})\right) P_{j_1,j_2,j_3}(\eta), \tag{5.13}$$

$$E_0^X\left(\prod_{k=1}^{4} q_{j_k}^{\varepsilon_k}\right) = \frac{h^4}{16}\exp\left(b^2\sum_{1\le s<r\le 4}\eta_s\eta_r(X_{j_s}, X_{j_r})\right) P_{j_1,j_2,j_3,j_4}(\eta). \tag{5.14}$$

5.3.4 Evaluation of probabilities $P_{j_1,\dots,j_m}(\eta)$

By definition of $(z_1,\dots,z_m)$ we have

$$E_0^X z_k = 0, \quad E_0^X z_k^2 = 1, \quad E_0^X z_k z_s \triangleq r_{ks}(\eta) = \frac{\eta_k\eta_s(X_{j_k}, X_{j_s})}{\|X_{j_k}\|\|X_{j_s}\|}, \quad 1\le k<s\le m.$$

Denote $\tilde T_{j_k} = T_{j_k} - a_{j_k}$. Observe that

$$\begin{aligned} P_{j_1,\dots,j_m}(\eta) &= 1 - \sum_{k=1}^{m}\Phi(-\tilde T_{j_k} - bm_{j_k}(\eta)) \\ &+ O\left(\sum_{1\le k<s\le m} P_{r_{ks}(\eta)}\left(-\tilde T_{j_k} - bm_{j_k}(\eta), -\tilde T_{j_s} - bm_{j_s}(\eta)\right)\right), \end{aligned} \tag{5.15}$$

where we set, for the Gaussian random vector (z_1, z_2) with $Ez_k = 0, Ez_k^2 = 1, k = 1,2, Ez_1z_2 = r$,

$$P_r(t_1,t_2) = P(z_1 < t_1, z_2 < t_2) = P(z_1 > -t_1, z_2 > -t_2).$$

The control of $P_{j_1,\dots,j_m}(\eta)$ then depends on the sequence x_{np}.
CASE 1: $x = 0$. Under the event $\mathcal{X}_{n,p}$, we have $\max_j a_j = o(\sqrt{\log(p)})$ and $\tilde T_{j_k}/\sqrt{\log(p)} \to \infty$. Under the event $\mathcal{X}_{np}$, we have

$$b|m_{jk}(\eta)| \le b\sum_{s\ne k}|(X_{j_s}, X_{j_k})|/\|X_{j_k}\| \le o(b\sqrt{n}/\log(p)) = o(1/\sqrt{\log(p)}).$$

It follows that

$$\max_j \Phi(-\tilde T_{j_k} - bm_{j_k}(\eta)) = o(p^{-\alpha}), \quad \forall\ \alpha > 0.$$

We conclude

$$P_{j_1,\ldots,j_m}(\eta) = 1 - O\left(\sum_{k=1}^{m} \Phi(-\tilde{T}_{j_k} - bm_{j_k}(\eta))\right) = 1 - o(p^{-\alpha}), \quad \forall\ \alpha > 0. \quad (5.16)$$

CASE 2: $x > 0$. We have under the event $\mathcal{X}_{n,p}$, $b|m_{jk}(\eta)| = o(b\sqrt{n}/\log(p)) = o(1)$ and $\tilde{T}_{jk}b = O(\log(p)/\sqrt{n})$. Hence, $\tilde{T}_{jk}b|m_{jk}(\eta)| = o(1)$. Applying Lemma 7.2, we bound the first term in (5.15)

$$\begin{aligned}\sum_{k=1}^{m} \Phi(-\tilde{T}_{j_k} - bm_{j_k}(\eta)) &= \sum_{k=1}^{m} \Phi(-\tilde{T}_{j_k}) - b\sum_{k=1}^{m} m_{j_k}(\eta)\Phi(-\tilde{T}_{j_k}) \\ &+ b^2 \sum_{k=1}^{m} O\left(m_{j_k}^2(\eta)\tilde{T}_{j_k}\Phi(-\tilde{T}_{j_k})\right) .\end{aligned}$$

Let us define

$$R_m \triangleq \sum_{k=1}^{m} \Phi(-\tilde{T}_{j_k}) = o((ph)^{-1}), \quad (5.17)$$

by (7.1) and (7.2) since $x < \varphi(\beta) \leq \sqrt{2}(1 - \sqrt{1-\beta})$. Applying again (7.1) and (7.2), we get

$$\begin{aligned}b(X_{j_s}, X_{j_k})\Phi(-\tilde{T}_{j_k})/\|X_{j_k}\| &= O\left[T_{j_k}(X_{j_s}, X_{j_k})\Phi(-\tilde{T}_{j_k})n^{-1}\right] \\ &= O\left[T_{j_k}\Phi(-\tilde{T}_{j_k})|r_{ks}|\right] = o(|r_{ks}|/(ph)) \\ b^2(X_{j_s}, X_{j_k})^2\tilde{T}_{j_k}\Phi(-\tilde{T}_{j_k})n^{-1} &= O(T_{j_k}^3)\Phi(-\tilde{T}_{j_k})r_{ks}^2 = o(r_{ks}^2/(ph)).\end{aligned}$$

It follows that

$$\begin{aligned}\sum_{k=1}^{m} \Phi(-\tilde{T}_{j_k} - bm_{j_k}(\eta)) &= R_m - \sum_{1\leq k<s\leq m} o\left(|r_{ks}|/(ph)\right) + \sum_{1\leq k<s\leq m} o\left(r_{ks}^2/(ph)\right) \\ &= \sum_{1\leq k<s\leq m} o\left((1 + |r_{ks}| + r_{ks}^2)/(ph)\right) .\end{aligned}$$

Let us turn to the second term in (5.15). If $\tilde{T}_{j_k} \geq \log(p)$, then

$$P_{r_{ks}(\eta)}\left(-\tilde{T}_{j_k} - bm_{j_k}(\eta), -\tilde{T}_{j_s} - bm_{j_s}(\eta)\right) \leq \Phi(-\tilde{T}_{j_k} - bm_{j_k}(\eta)) = o(\left((ph)^{-2}\right)$$

If $\tilde{T}_{j_k} \leq \log(p)$, we have $\tilde{T}_{j_k}r_{ks} = o(1)$ under the event $\mathcal{X}_{np}$. By Lemma 7.3 and previous evaluations, we get

$$\begin{aligned}&P_{r_{ks}(\eta)}\left(-\tilde{T}_{j_k} - bm_{j_k}(\eta), -\tilde{T}_{j_s} - bm_{j_s}(\eta)\right) \\ &= \Phi(-\tilde{T}_{j_k} - bm_{j_k}(\eta))\,\Phi(-\tilde{T}_{j_s} - bm_{j_s}(\eta))O\left(1 + r_{ks}^2 + |r_{ks}|\right) = o\left((ph)^{-2}\right).\end{aligned}$$

Finally, we obtain

$$\begin{aligned}P_{j_1,\ldots,j_m}(\eta) &= 1 - R_m + o\left(\sum_{1\leq k<s\leq m} |r_{ks}|/ph\right) + o((ph)^{-2}) \quad (5.18) \\ &= 1 + o((ph)^{-1}). \quad (5.19)\end{aligned}$$

5.3.5 Evaluation of A_2

We have $b^2 \max_{1\le j_1<j_2\le p} |(X_{j_1}, X_{j_2})| = o(1)$ under the event $\mathcal{X}_{n,p}$. Since $P_{j_1,j_2}(\eta) = O(1)$, we get from (5.12)

$$E_0^X \left(\prod_{k=1}^{2} q_{j_k}^{\varepsilon_k} \right) = O(h^2).$$

By Assumption **B1**, we have

$$\sup_{j_1 \neq j_2} E_X (X_{j_1}, X_{j_2})^2 = O(n) .$$

It then follows from (5.6) and (5.9) that A_2 is of the order

$$b^4 h^2 \sum_{1\le j_1<j_2\le p} (X_{j_1}, X_{j_2})^2 \asymp p^2 h^2 b^4 n \asymp nk^2 b^4 \to 0 ,$$

in P_X-probability.

5.3.6 Evaluation of A_3

Let us turn to A_3. Consider η_k as independent random variables taking values in $\{-1, 1\}$ with probabilities 1/2. By (5.10) and (5.13), we can write

$$E_{\tilde{\pi}_Z} \left(\theta_{j_1}^2 \theta_{j_2} \theta_{j_3} \right) = \frac{h^3}{8} E_\eta \left(\eta_2 \eta_3 \exp \left(b^2 \sum_{1\le s<r\le 3} \eta_s \eta_r (X_{j_s}, X_{j_r}) \right) P_{j_1,j_2,j_3}(\eta) \right) .$$

Under the event $\mathcal{X}_{n,p}$ it follows from (5.16), (5.19), and the definition of $\mathcal{X}_{n,p}$ that

$$P_{j_1,j_2,j_3}(\eta) = 1 + o((ph)^{-1}) \text{ and } b^2 \sum_{1\le s<r\le 3} \eta_s \eta_r (X_{j_s}, X_{j_r}) = o(1).$$

It follows that

$$E_{\tilde{\pi}_Z} \left(\theta_{j_1}^2 \theta_{j_2} \theta_{j_3} \right) \;=\; \frac{h^3}{8} E_\eta \left(\eta_2 \eta_3 \exp \left(b^2 \sum_{1\le s<r\le 3} \eta_s \eta_r (X_{j_s}, X_{j_r}) \right) \right) + o\left(h^2 p^{-1} \right) .$$

By Taylor expansion of the exponential function, the expectation over η is of the form, for $c_{sr} = b^2 (X_{j_s}, X_{j_r})$,

$$\begin{aligned} &E_\eta \left(\eta_2 \eta_3 \left(1 + \eta_1 \eta_2 c_{12} + \eta_1 \eta_3 c_{13} + \eta_2 \eta_3 c_{23} + O\left(c_{12}^2 + c_{13}^2 + c_{23}^2 \right) \right) \right) \\ &\qquad = b^2 (X_{j_2}, X_{j_3}) + O\left(b^4 \sum_{1\le s<r\le 3} (X_{j_s}, X_{j_r})^2 \right) . \end{aligned}$$

Under the event $\mathcal{X}_{n,p}$, we derive from (5.7) that

$$A_3 \le h^3 \left(b^6 O(H_1) + b^8 O(H_2) \right) + b^4 o(H_3 h^2 p^{-1}),$$

where

$$\begin{aligned}
H_1 &= \sum_{1\le j_1<j_2<j_3\le p} |(X_{j_1},X_{j_2})||(X_{j_1},X_{j_3})||(X_{j_2},X_{j_3})|,\\
H_2 &= \sum_{1\le j_1<j_2<j_3\le p}\sum_{1\le s<r\le 3} |(X_{j_1},X_{j_2})||(X_{j_1},X_{j_3})|(X_{j_s},X_{j_r})^2,\\
H_3 &= \sum_{1\le j_1<j_2<j_3\le p} |(X_{j_1},X_{j_2})||(X_{j_1},X_{j_3})|.
\end{aligned}$$

Since

$$\begin{aligned}
|(X_{j_1},X_{j_2})||(X_{j_1},X_{j_3})||(X_{j_r},X_{j_s})| &\le |(X_{j_1},X_{j_2})|^3+|(X_{j_1},X_{j_3})|^3+|(X_{j_r},X_{j_s})|^3,\\
|(X_{j_1},X_{j_2})||(X_{j_1},X_{j_3})|(X_{j_2},X_{j_3})^2 &\le (X_{j_1},X_{j_2})^4+(X_{j_1},X_{j_3})^4+(X_{j_2},X_{j_3})^4,\\
|(X_{j_1},X_{j_2})||(X_{j_1},X_{j_3})| &\le (X_{j_1},X_{j_2})^2+(X_{j_1},X_{j_3})^2,
\end{aligned}$$

we derive from (3.6)

$$E_X H_1 = O(p^3 n^{3/2}), \quad E_X H_2 = O(p^3n^2), \quad E_X H_1 = O(p^3 n).$$

Applying Markov's inequality yields

$$H_1 = O_{P_X}(p^3n^{3/2}), \quad H_2 = O_{P_X}(p^3n^2), \quad H_1 = O_{P_X}(p^3n).$$

Combining these bounds, we obtain

$$A_3 = O_{P_X}((b^6h^3p^3n^{3/2}) + O_{P_X}(b^8h^3p^3n^2) + b^4 o_{P_X}(b^4h^2p^2n).$$

Since $b^4k^2n = o(1),\ hp \asymp k,\ b = o(1)$, we get $A_3 = o_{P_X}(1)$.

5.3.7 Evaluation of A_4

Let us evaluate the item A_4. Similarly to A_3, we can write

$$E_{\tilde{\pi}_Z}\left(\theta_{j_1}\theta_{j_2}\theta_{j_3}\theta_{j_4}\right) = h^4 E_\eta\left(\eta_1\eta_2\eta_3\eta_4 \exp\left(b^2\sum_{1\le s<r\le 4}\eta_s\eta_r(X_{j_s},X_{j_r})\right)P_{j_1,j_2,j_3,j_4}(\eta)\right) \tag{5.20}$$

Under the event $\mathcal{X}_{n,p}$ we have

$$b^2\sum_{1\le s<r\le 4}\eta_s\eta_r(X_{j_s},X_{j_r}) = o(1).$$

CASE 1: $x>0$. By (5.18), we have

$$P_{j_1,j_2,j_3,j_4}(\eta) = 1 - R_4 + o\left(\sum_{1\le s<r\le 4}|r_{sr}|/hp\right) + o((ph)^{-2}).$$

Applying a Taylor expansion of the exponential term in (5.20) yields

$$\begin{aligned}
& E_\eta\left(\eta_1\eta_2\eta_3\eta_4 \exp\left(b^2 \sum_{1\le s<r\le 4} \eta_s\eta_r (X_{j_s}, X_{j_r})\right) P_{j_1,j_2,j_3,j_4}(\eta)\right) \\
&= E_\eta\left(\eta_1\eta_2\eta_3\eta_4\left(1 + b^2 \sum_{1\le s<r\le 4} \eta_s\eta_r (X_{j_s}, X_{j_r})\right)(1-R_4)\right) + O(|\delta_1|) + O(|\delta_2|) \\
&= O(|\delta_1|) + O(|\delta_2|),
\end{aligned}$$

where

$$\begin{aligned}
\delta_1 &= O\left(b^4 \sum_{1\le s<r\le 4} (X_{j_s}, X_{j_r})^2\right), \\
\delta_2 &= o\left(\sum_{1\le k<s\le 4} |r_{ks}|/ph\right) + o((ph)^{-2})\ .
\end{aligned}$$

CASE 2: $x = 0$. By (5.16), $P_{j_1,j_2,j_3,j_4}(\eta) = 1 - o(p^{-2})$. Arguing as in Case 1, we get

$$\begin{aligned}
& E_\eta\left(\eta_1\eta_2\eta_3\eta_4 \exp\left(b^2 \sum_{1\le s<r\le 4} \eta_s\eta_r (X_{j_s}, X_{j_r})\right) P_{j_1,j_2,j_3,j_4}(\eta)\right) \\
&= O\left(b^4 \sum_{1\le s<r\le 4} (X_{j_s}, X_{j_r})^2\right) + o(p^{-2}).
\end{aligned}$$

All in all, we obtain that under the event $\mathcal{X}_{n,p}$,

$$A_4 \le h^4 b^8 O(H_1) + \begin{cases} o(H_2 b^4 h^4/p^2), & x = 0, \\ o(H_3 b^4 h^3/np + H_2 b^4 h^2/p^2), & x > 0, \end{cases}$$

where

$$\begin{aligned}
H_1 &= \sum_{1\le j_1<j_2<j_3<j_4\le p} |(X_{j_1}, X_{j_2})||(X_{j_3}, X_{j_4})| \sum_{1\le s<r\le 4} (X_{j_s}, X_{j_r})^2, \\
H_2 &= \sum_{1\le j_1<j_2<j_3<j_4\le p} |(X_{j_1}, X_{j_2})||(X_{j_3}, X_{j_4})|, \\
H_3 &= \sum_{1\le j_1<j_2<j_3<j_4\le p} |(X_{j_1}, X_{j_2})||(X_{j_3}, X_{j_4})| \sum_{1\le s<r\le 4} |(X_{j_s}, X_{j_r})|.
\end{aligned}$$

We combine the classical upper bounds,

$$\begin{aligned}
|(X_{j_1}, X_{j_2})||(X_{j_3}, X_{j_4})|(X_{j_s}, X_{j_r})^2 &\le (X_{j_1}, X_{j_2})^4 + (X_{j_3}, X_{j_4})^4 + (X_{j_s}, X_{j_r})^4, \\
|(X_{j_1}, X_{j_2})||(X_{j_3}, X_{j_4})| &\le (X_{j_1}, X_{j_2})^2 + (X_{j_3}, X_{j_4})^2, \\
|(X_{j_1}, X_{j_2})||(X_{j_3}, X_{j_4})||(X_{j_s}, X_{j_r})| &\le |(X_{j_1}, X_{j_2})|^3 + |(X_{j_3}, X_{j_4})|^3 + |(X_{j_s}, X_{j_r})|^3\ .
\end{aligned}$$

with (3.6) and obtain

$$E_X(H_1) = O(p^4n^2), \quad E_X(H_2) = O(p^4n), \quad E_X(H_3) = O(p^4n^{3/2}).$$

Applying Markov's inequality yields

$$H_1 = O_{P_X}(p^4n^2),\ H_2 = O_{P_X}(p^4n),\ H_3 = O_{P_X}(p^4n^{3/2}).$$

Since $b^4k^2n = o(1),\ hp \asymp k$, we get

$$h^4b^8H_1 = O(h^4p^4b^8n^2) = o_{P_X}(1), \quad H_2b^4h^2/p^2 = O_{P_X}(b^4h^2p^2n) = o_{P_X}(1).$$

If $x > 0$, we also have to upper bound the term H_3. Since $r^2_{n,p} = o(1/\sqrt{n})$ (cf. (4.1)) and since $x > 0$, we derive that $k = o(\sqrt{n})$. Then, we get

$$H_3b^4h^3/np = O_{P_X}(b^6p^3h^3n^{3/2}/nb^2) = o_{P_X}(1),$$

Therefore we obtain $A_4 = o_{P_X}(1)$. The proposition follows. $\square$

5.4 Proof of Proposition 5.2

We will prove that there exists a family of events $\mathcal{Z}_{n,p}$ such that $P_0(\mathcal{Z}_{n,p}) \to 1$ and

$$\log(\Lambda(Z)) = \sum_{j=1}^{p} \log\left(1 + (h/2)\left(e^{d_j^+} + e^{d_j^-} - 2\right)\right) \to 0, \quad Z \in \mathcal{Z}_{n,p}.$$

We take $\mathcal{Z}_{n,p} = \{(X,Y) : |y'_j| \le T_j,\ 1 \le j \le p,\ X \in \mathcal{X}_{n,p}\}$ where $\mathcal{X}_{n,p}$ was defined in Section 5.3.2. It follows from Lemma 5.2 and Section 5.3.2 that $P_0(\mathcal{Z}_{n,p}) \to 1$.

Under the events $\mathcal{Z}_{n,p}$ we can replace the quantities $(h/2)e^{d_j^\pm}/2$ by $q_j^\pm = (h/2)e^{d_j^\pm}\mathbb{1}_{\pm y'_j < T_j}$, cf. Section 5.3.1. Let us consider

$$\tilde{L} = \sum_{j=1}^{p} \log(1 + \Delta_j), \quad \Delta_j = (q_j^+ + q_j^- - h).$$

Under the event $\mathcal{A} = \mathcal{A}_{n,p} = \bigcap_{j=1}^p(\mathcal{A}_j^+ + \mathcal{A}_j^-)$ defined in Section 5.3.1, we have uniformly in $1 \le j \le p$,

$$\begin{aligned} q_j^+ + q_j^- &= \frac{h}{2}e^{-a_j^2/2}\cosh(a_jy'_j) \le he^{-a_j^2/2}\cosh(a_jT_j) \\ &= (1 + e^{-2a_jT_j})/2 \sim 1/2, \end{aligned}$$

as $h \to 0$. Consequently, we have

$$\tilde{L} = \sum_{j=1}^{p} = A_1 + O(A_2), \quad A_1 = \sum_{j=1}^{p}\Delta_j, \quad A_2 = \sum_{j=1}^{p}\Delta_j^2.$$

Thus, we need to show that $A_1 \to 0$ and that $A_2 \to 0$ in P_0-probability. It was stated in the proof of Lemma 5.3 that $E_0^X A_2 = o_{P_X}(1)$. Markov's inequality then allows to derive that $A_2 = o_{P_X}(1)$. In order to prove the first relation, we shall show that $E_0^X A_1 \to 0$ and that $\mathrm{Var}_0^X A_1 \to 0$ in P_0-probability. Observe that

$$E_0^X A_1 = h \sum_{j=1}^{p} (\Phi(T_j - a_j) - 1) = -h \sum_{j=1}^{p} \Phi(-T_j + a_j).$$

By (7.1) and (7.2) we have

$$h \sum_{j=1}^{p} \Phi(-T_j + a_j) \asymp \sum_{j=1}^{p} \Phi(-T_j) = o(1).$$

We have $\mathrm{Var}_0^X A_1 \le B + A_2$ with $B = \sum_{1 \le j < l \le p} \hat{\Delta}_j \hat{\Delta}_l$ and $\hat{\Delta}_j = \Delta_j - E_0^X \Delta_j$. We need to check that, in P_X probability,

$$E_0^X(B) = \sum_{1 \le j < l \le p} E_0^X(\hat{\Delta}_j \hat{\Delta}_l) \to 0.$$

Note that

$$E_0^X(\hat{\Delta}_j \hat{\Delta}_l) = B_{jl} - C_{jl},$$

where

$$B_{jl} = E_0^X \left((q_j^+ + q_j^-)(q_l^+ + q_l^-) \right), \quad C_{jl} = h^2 \Phi(T_j - a_j) \Phi(T_l - a_l) \ .$$

We consider independent random variables η_1, η_2 taking values -1 and 1 with probabilities $1/2$. We write (compare with (5.12))

$$B_{jl} = h^2 E_\eta \left[\exp\left(\eta_1 \eta_2 b^2(X_j, X_l) \right) P_{j,l}(\eta) \right] \ , \quad C_{jl} = h^2 P_{j,l}^0 \ .$$

Here we set

$$P_{j,l}^0 = \Phi(\tilde{T}_j)\Phi(\tilde{T}_l) = 1 - \Phi(-\tilde{T}_j) - \Phi(-\tilde{T}_l) + \Phi(-\tilde{T}_j)\Phi(-\tilde{T}_l), \ \tilde{T}_l = T_l - a_l.$$

We obtain the new decomposition

$$B_{jl} - C_{jl} = h^2(U_{jl} + V_{jl}) \ , \tag{5.21}$$

where

$$U_{jl} = E_\eta \left[\left(\exp\left((\eta_1 \eta_2 b^2(X_j, X_l) \right) - 1 \right) P_{j,l}(\eta) \right], \quad V_{jl} = E_\eta \left(P_{j,l}(\eta) - P_{j,l}^0 \right) \ .$$

Let us recall some notations introduced in Section 5.3.4. $r_{jl}(\eta) = \eta_1 \eta_2 r_{jl}$,

$$r_{jl} = \frac{(X_j, X_l)}{\|X_j\| \|X_l\|}, \quad m_{jl}(\eta) = \frac{\eta_1 \eta_2 (X_j, X_l)}{\|X_j\|}, \quad m_{lj}(\eta) = \frac{\eta_1 \eta_2 (X_j, X_l)}{\|X_l\|} \ .$$

Moreover, z_j and z_l stand for standard Gaussian variables with $\mathrm{Cov}(z_j, z_l) = r_{jl}(\eta)$. Then, $P_{j,l}(\eta)$ is written as

$$\begin{aligned} P_{j,l}(\eta) &= P_0^X(z_j < \tilde{T}_j - bm_{jl}(\eta),\, z_l < \tilde{T}_l - bm_{lj}(\eta)) \\ &= 1 - \Phi(-\tilde{T}_j + bm_{jl}(\eta)) - \Phi(-\tilde{T}_l + bm_{lj}(\eta)) \\ &\quad + P_0^X(z_j < -\tilde{T}_j + bm_{jl}(\eta),\, z_l < -\tilde{T}_l + bm_{lj}(\eta)) \ . \end{aligned}$$

CASE 1: $x = 0$. The evaluations of the terms V_{jl} in (5.21) are similar to the ones in Section 5.3.4. We get

$$P^0_{j,l} = 1 - o(p^{-2}), \quad P_{j,l}(\eta) = 1 - o(p^{-2}), \quad |P_{j,l}(\eta) - P^0_{j,l}| = o(p^{-2}).$$

We derive that $h^2 \sum_{1\le j<l\le p} V_{j,l} = o(h^2)$.

CASE 2: $x > 0$. We have (compare with (5.17) and (5.19))

$$\begin{aligned} \Phi(-\tilde{T}_j)\Phi(-\tilde{T}_l) &= o((ph)^{-2}), \\ P_0^X(z_j < -\tilde{T}_j + bm_{jl}(\eta),\, z_l < -\tilde{T}_l + bm_{lj}(\eta)) &= o((ph)^{-2}), \\ \Phi(-\tilde{T}_j + bm_{jl}(\eta)) &= \Phi(-\tilde{T}_j) + \eta_1\eta_2 b r_{jl} + o(r^2_{jl}/(ph)) \ , \end{aligned}$$

Taking the expectation over η, we get

$$E_\eta\left(P_{j,l}(\eta)\right) - P^0_{j,l} = o(r^2_{jl}/(hp)) + o((ph)^{-2}).$$

in P_X-probability. Therefore

$$h^2 \sum_{1\le j<l\le p} V_{jl} = O(Hhp^{-1}) + o(1), \quad H = \sum_{1\le j<l\le p} r^2_{jl}.$$

Under $\mathcal{X}_{n,p}$ we have $r^2_{jl} \sim n^{-2}(X_j, X_l)^2$. Since $E_X[(X_j, X_l)^2] = O(n)$ for $j \neq l$ (Assumption **B1**), we get

$$H \sim n^{-2} \sum_{1\le j<l\le p} (X_j, X_l)^2 = O_{P_X}(n^{-1}p^2) \ .$$

This leads to $h^2 \sum_{1\le j<l\le p} V_{jl} = O_{P_X}(ph/n) + o(1)$. Since $r^2_{n,p} = o(1/\sqrt{n})$ (Eq. 4.1) and since $x > 0$, we derive that $k = o(\sqrt{n})$. Consequently, we have $ph/n = O(k/n) = o(1)$.

Let us turn to the terms U_{jl}. They are handled as in Section 5.3.2. We have

$$\begin{aligned} U_{jl} &= E_\eta\left(\left(\eta_1\eta_2 b^2(X_j, X_l) + O\left(b^4(X_j, X_l)^2\right)\left(1 + o((ph)^{-1}\right)\right)\right) \\ &= O\left(b^4(X_j, X_l)^2\right) + O\left(b^2|(X_j, X_l)|/(ph)\right). \end{aligned}$$

Then, we get

$$h^2 \sum_{1\le j<l\le p} U_{jl} = O\left(h^2 b^4 H_1\right) + O\left(hb^2 H_2/p\right),$$

where

$$H_1 = \sum_{1\le j<l\le p} (X_j, X_l)^2, \quad H_2 = \sum_{1\le j<l\le p} |(X_j, X_l)| \le pH_1^{1/2}.$$

Arguing as for H, we get

$$H_1 = O_{P_X}(p^2 n), \quad H_2 = O_{P_X}(p^2 n^{1/2}).$$

It follows that

$$\sum_{1\le j<l\le p} B_{jl} = O_{P_X}(p^2h^2b^4n) + O_{P_X}(phb^2n^{1/2}) = o_{P_X}(1),$$

since $p^2h^2b^4n \asymp k^2b^4n \to 0$ by (4.1). The proposition follows. $\square$

5.5 Proof of Theorem 4.5

An in the proof of Theorem 4.1, we consider $x = \limsup x_{n,p}$ and we take $c \in (0,1)$ such that $x_c = x/c < \varphi(\beta)$. We also define $b = x_c\sqrt{\log(p)/n}$. We first consider the case where $k\log(p)/n \to 0$.

We use a different prior π than for Theorem 4.1. Let us note $\mathcal{M}(k,p)$ the collection of subsets of $\{1,\ldots,p\}$ of size k. We consider a random vector $\theta = (\theta_j)$ with coordinates $\theta_j = b\epsilon_j$ where $\epsilon_j \in (0,1)$. The set of non-zero coefficient of ϵ is drawn uniformly in $\mathcal{M}(k,p)$. This introduces a prior probability π on θ.

Consider the mixture

$$P_\pi(dZ) = E_\pi P_{\theta,\sqrt{1-bk^2}}(dZ) = \int_{\mathbb{R}^p} P_{\theta,\sqrt{1-bk^2}}(dZ)\pi(d\theta)$$

and the likelihood ratio

$$L_\pi(Z) = \frac{dP_\pi}{dP_{0,1}}(Z)\ .$$

As in the proof of Theorem 4.1, we shall prove that $L_\pi(Z)$ converges to 1 in P_0 probability. This will enforce that $\gamma^{un}_{n,p,k}[x_c\sqrt{k\log(p)}/\sqrt{1-kb^2}] \to 1$. Since kb^2 converges to 0, this will complete the proof.

The likelihood ratio has the form $L_\pi(Z) = \sum_{m\in\mathcal{M}(k,p)} |\mathcal{M}(k,p)|^{-1}L_m(Z)$ and

$$\begin{aligned} L_m(Z) &= (1-kb^2)^{-n/2}\exp\left(-\frac{kb^2\|Y\|^2}{2(1-kb^2)} + \frac{b(Y,\sum_{i\in m}X_i)}{1-kb^2}\right) \\ &\times \exp\left[-\sum_{i,j\in m}\frac{b^2}{2(1-kb^2)}(X_i,X_j)\right]\ . \end{aligned} \tag{5.22}$$

Definition 5.1 *Consider $\delta\in(0,1)$, a positive integer s and a $n\times p$ matrix A. We say that A satisfies a δ-restricted isometry property of order s if for all $\theta\in\mathbb{R}^p_s$,*

$$(1-\delta)\|\theta\| \le \|A\theta\| \le (1+\delta)\|\theta\|\ .$$

Let us define the events Ω_1 and Ω_2 by

$$\Omega_1 : \text{"}X/\sqrt{n} \text{ satisfies a } \delta_{n,p}^{(1)} \text{ restricted isometry of order } 2k\text{"}$$

$$\Omega_2 : \text{"For any } 1 \leq i \leq p,\ (Y, X_i/\|X_i\|) \leq \sqrt{2\log(p)}(1 + \delta_{n,p}^{(2)})\text{"}\ ,$$

where $\delta_{n,p}^{(1)} = 16\sqrt{k\log(p)/n}$ and $\delta_{n,p}^{(2)} = \log^{-1/2}(p)$. Applying a deviation inequality due to Davidson and Szarek (Theorem 2.13 in [6]), we derive that $P_X(\Omega_1^c) = o(1)$. By the Gaussian concentration inequality, we have $P_0(\Omega_2^c) = o(1)$. Then, we take $\Omega = \Omega_1 \cap \Omega_2$.

Lemma 5.4 *We have* $E_0\left[L_\pi^2(Z)\mathbb{1}_\Omega\right] \leq 1 + o(1)$.

Lemma 5.5 *We have* $E_0\left[L_\pi(Z)\mathbb{1}_{\Omega^c}\right] = o(1)$.

Since $E_0\left[L_\pi(Z)\right] = 1$, we get the desired result by combining these two lemmas.

Let us turn to the case $k\log(p)/n \to \infty$. We consider $b > 0$ defined by

$$\frac{kb^2}{1 - kb^2} = (2\beta - 1)\frac{k\log(p)}{n}\ .$$

Lemma 5.6 *We have*

$$E_0\left[L_\pi^2(Z)\right] = 1 + o(1)\ .$$

This lemma implies that for $r = \sqrt{(2\beta - 1)k\log(p)/n} \to \infty$, we have $\gamma_{n,p,k}^{un}(r) \to 1$. $\square$

In the proof of the following lemmas, $o(1)$ stands for a positive quantity which depends only on (k, p, n) and tends to 0 as (n, p) tend to infinity.

5.5.1 Proof of Lemma 5.4

In order to upper bound $E_0\left[L_\pi^2(Z)\mathbb{1}_\Omega\right]$, we first upper bound $E_0\left[L_{m_1}(Z)L_{m_2}(Z)\mathbb{1}_\Omega\right]$ for any $m_1, m_2 \in \mathcal{M}(k, p)$. We define W_1, W_2, W_3 by , $W_1 = \sum_{i\in m_1\setminus m_2} X_i$, $W_2 = \sum_{i\in m_2\setminus m_1} X_i$, and $W_3 = \sum_{i\in m_1\cap m_2} X_i$. We note $S = |m_1 \cap m_2|$.

$$\begin{aligned} L_{m_1}(Z)L_{m_2}(Z) &= (1 - kb^2)^{-n}\exp\left(-\frac{kb^2\|Y\|^2}{1 - kb^2} + \frac{b(Y, 2W_3 + W_1 + W_2)}{1 - kb^2}\right) \\ &\times \exp\left[-\frac{b^2}{2(1 - kb^2)}(\|W_1 + W_3\|^2 + \|W_2 + W_3\|^2)\right]\ . \end{aligned}$$

Let us take the expectation of $L_{m_1}(Z)L_{m_2}(Z)$ with respect to (W_1, W_2).

$$E_0^{Y,W_3}[L_{m_1}(Z)L_{m_2}(Z)] = (1 - Sb^2)^{-n}\exp\left[-\frac{\|Y\|^2 Sb^2}{1 - Sb^2} + \frac{2b(Y, W_3)}{1 - Sb^2} - \frac{b^2\|W_3\|^2}{1 - Sb^2}\right]\ .$$

When $S=0$, we have $E_0^{Y,W_3}[L_{m_1}(Z)L_{m_2}(Z)]=1$. Let us now consider the case $S>0$. On the event Ω, we have

$$(Y,\frac{W_3}{\|W_3\|})\leq\sqrt{2\log(p)}(1+\delta^{(2)}_{n,p})\frac{\sum_{i\in m_1\cap m_2}\|X_i\|}{\|\sum_{i\in m_1\cap m_2}X_i\|}\leq\sqrt{2S\log(p)}(1+o(1))\ ,$$

since $X/\sqrt{n}$ satisfies a $\delta^{(1)}_{n,p}$-restricted isometry of order $2k$. Then, we can upper bound the expectation with respect to Y.

$$\begin{aligned}E_0^{W_3}[\mathbb{1}_\Omega L_{m_1}(Z)L_{m_2}(Z)] &\leq (1-S^2b^4)^{-n/2}\exp\left[\frac{b^2\|W_3\|^2}{1+Sb^2}\right]\\ &\times\ \Phi\left[\sqrt{2S\log(p)}(1+o(1))-2b\|W_3\|(1-o(1))\right]\ .\end{aligned}$$

Moreover on Ω, we have $\sqrt{1-\delta^{(1)}_{n,p}}\leq\|W_3\|/\sqrt{nS}\leq\sqrt{1+\delta^{(1)}_{n,p}}$. Since $k\log(p)/n$ goes to 0, we get

$$E_0[\mathbb{1}_\Omega L_{m_1}(Z)L_{m_2}(Z)]\ \leq\ \exp\left[x_c^2S\log(p)(1+o(1))\right]\Phi\left[\sqrt{S\log(p)}\left(\sqrt{2}-2x_c+o(1)\right)\right]\ .$$

For any $x<0$, we have $\Phi(x)\leq e^{-x^2/2}$. Hence, we get $\Phi(x)\leq e^{-x_-^2/2}$ for any $x\in\mathbb{R}$. It follows that

$$E_0[\mathbb{1}_\Omega L_{m_1}(Z)L_{m_2}(Z)]\ \leq\ \exp\left[S\log(p)\left\{x_c^2-(1-\sqrt{2}x_c)_-^2+o(1)\right\}\right]\ . \quad (5.23)$$

Hence, we get

$$E_0[\mathbb{1}_\Omega L_\pi^2(Z)]\leq E_S\left[p^{S\{x_c^2-(1-\sqrt{2}x_c)_-^2+o(1)\}}\right]$$

where S follows a hypergeometric distribution with parameters p, k and k/p. We know from Aldous (p.173) [1] that S has the same distribution as the random variable $E(U|\mathcal{B}_p)$ where U is binomial random variable of parameters k, k/p and $\mathcal{B}_p$ some suitable σ-algebra. By a convexity argument, we then obtain

$$\begin{aligned}E_0[\mathbb{1}_\Omega L_\pi^2(Z)] &\leq \left[1+\frac{k}{p}\left(p^{x_c^2-(1-\sqrt{2}x_c)_-^2+o(1)}-1\right)\right]^k\\ &\leq \exp\left[\frac{k^2}{p}p^{x_c^2-(1-\sqrt{2}x_c)_-^2+o(1)}\right]\\ &\leq \exp\left[p^{1-2\beta+x_c^2-(1-\sqrt{2}x_c)_-^2+o(1)}\right]\end{aligned}$$

Since $x_c<\varphi(\beta)$, one can check that $1-2\beta+x_c^2-(1-\sqrt{2}x_c)_-^2$ is negative and we conclude that $E_0[\mathbb{1}_\Omega L_\pi^2(Z)]\leq 1+o(1)$. $\square$

5.5.2 Proof of Lemma 5.5

By symmetry, it is sufficient to prove that $E_0(L_m(Z)\mathbb{1}_{\Omega^c}) = o(1)$. Let us decompose $E_0(L_m(Z)\mathbb{1}_{\Omega^c}) = E_0(L_m(Z)\mathbb{1}_{\Omega_2^c}) + E_0(L_m(Z)\mathbb{1}_{\Omega_1^c\cup\Omega_2})$. Since $E_0^X(L_m(Z)) = 1$, P_X almost surely, we have $E_0(L_m(Z)\mathbb{1}_{\Omega_2^c}) = P_X(\Omega_2^c) = o(1)$. Let us turn to $E_0(L_m(Z)\mathbb{1}_{\Omega_1^c\cup\Omega_2})$. For any $1 \leq i \leq p$, we define the event $\Omega^{(i)}$ by $(Y, X_i/\|X_i\|) \geq \sqrt{2\log(p)}(1+\delta_{n,p}^{(2)})$.

$$E_0(L_m(Z)\mathbb{1}_{\Omega_1^c\cup\Omega_2}) \leq \sum_{i=1}^{p} E_0\left[L_m(Z)\mathbb{1}_{\Omega_2}\mathbb{1}_{\Omega^{(i)}}\right]$$

The value of these expectations depends on i through the property "$i \in m$" or "$i \notin m$". Let us assume for instance that $1 \in m$ and $2 \notin m$. Then, we get

$$E_0(L_m(Z)\mathbb{1}_{\Omega_1^c\cup\Omega_2}) \leq kE_0\left[L_m(Z)\mathbb{1}_{\Omega_2}\mathbb{1}_{\Omega^{(1)}}\right] + pE_0\left[L_m(Z)\mathbb{1}_{\Omega_2}\mathbb{1}_{\Omega^{(2)}}\right] \ . \tag{5.24}$$

First, we upper bound $E_0[L_m(Z)\mathbb{1}_{\Omega_2}\mathbb{1}_{\Omega^{(2)}}]$. Taking the expectation of $L_m(Z)$ with respect to $(X_i)_{i\in m}$ leads to $E_0^{Y,X_2}[L_m(Z)] = 1$. Hence, we get

$$E_0[L_m(Z)\mathbb{1}_{\Omega_2}\mathbb{1}_{\Omega^{(2)}}] \leq P_0(\Omega^{(2)}) \leq p^{-1}e^{-\sqrt{\log(p)}} = o(p^{-1}) \ . \tag{5.25}$$

Let turn to $E_0[L_m(Z)\mathbb{1}_{\Omega_2}\mathbb{1}_{\Omega^{(1)}}]$. We first take the expectation of $L_m(Z)$ conditionally to X_1 and Y:

$$E_0^{Y,X_1}[L_m(Z)] = (1-b^2)^{-n/2}\exp\left[-\frac{b^2\|Y\|^2}{2(1-b^2)} - \frac{b^2\|X_1\|^2}{2(1-b^2)} + \frac{(Y,X_1)b}{1-b^2}\right] \ .$$

Then, we take the expectation with respect to Y

$$E_0^{X_1}[L_m(Z)\mathbb{1}_{\Omega^{(1)}}] \leq 1 - \Phi\left[\sqrt{\frac{2\log(p)}{1-b^2}}(1+\delta_{n,p}^{(2)}) - \frac{\|X_1\|b}{\sqrt{1-b^2}}\right] \ .$$

Moreover, on Ω_2 we have $\|X_1\| \leq \sqrt{n}(1+o(1))$

$$E_0^{X_1}[L_m(Z)\mathbb{1}_{\Omega^{(1)}\cup\Omega_2}] \leq \Phi\left[\sqrt{\log(p)}(x_c - \sqrt{2} + o(1))\right] \leq C\exp\left[-\log(p)(\sqrt{2} - x_c - o(1))^2/2\right]$$

for (n,p) large enough, since $x_c < \varphi(\beta) < \sqrt{2}$.

$$kE_0^{X_1}[L_m(Z)\mathbb{1}_{\Omega^{(1)}\cup\Omega_2}] \leq p^{-(\sqrt{2}-x_c)^2/2+1-\beta+o(1)} = o(1) \ , \tag{5.26}$$

since $x_c < \sqrt{2}(1-\sqrt{1-\beta}) \leq \varphi(\beta)$. Combining (5.24), (5.25), and (5.26) completes the proof. $\square$

5.5.3 Proof of Lemma 5.6

Arguing as in the proof of Lemma 5.4, we get

$$E_0^{W_3}[L_{m_1}(Z)L_{m_2}(Z)] \quad = \quad (1-S^2b^4)^{-n/2}\exp\left[\frac{b^2\|W_3\|^2}{1+Sb^2}\right] .$$

Taking the expectation with respect to W_3 leads to

$$E_0[L_{m_1}(Z)L_{m_2}(Z)] \quad = \quad (1-Sb^2)^{-n/2} \leq \exp\left[\frac{nSb^2}{2(1-kb^2)}\right]$$

As in the proof of Lemma 5.4, we upper bound the term $E_0[L_\pi^2(Z)]$ by Jensen's inequality.

$$\begin{aligned} E_0[L_\pi^2(Z)] &\leq \left[1+\frac{k}{p}\left\{\exp\left(\frac{nb^2}{2(1-kb^2)}\right)-1\right\}\right]^k \leq \exp\left[\frac{k^2}{p}\exp\left(\frac{nb^2}{2(1-kb^2)}\right)\right] \\ &\leq \exp\left[p^{1-2\beta}\exp\left\{(\beta-1/2)\log(p)\right\}\right] = 1+o(1) , \end{aligned}$$

since b satisfies $kb^2/(1-kb^2) = (2\beta-1)k\log(p)/n$. □

6 Proofs of the upper bounds

6.1 Tests based on the statistic t_0

Recall that

$$t_0 = (2n)^{-1/2}\sum_{i=1}^n (Y_i^2-1).$$

Under H_0, the statistics $Y_i = \xi_i \sim \mathcal{N}(0,1)$ are i.i.d. This implies $E_0(t_0) = 1$, $\mathrm{Var}_0(t_0) = 1$. By the Central Limit Theorem, $t_0 \to \xi \sim \mathcal{N}(0,1)$ as $n\to\infty$ in P_0-probability. This yields Theorem 4.2 (i).

Let us consider the type II errors. We need to show that, if $nr^4 \to \infty$, then $\sup_{\theta\in\Theta_p(r)} P_\theta(t_0 \leq u_\alpha) \to 0$. We will prove that, uniformly over $\theta \in \Theta_p(r)$,

$$E_\theta t_0 \to \infty, \quad \mathrm{Var}_\theta t_0 = o((E_\theta t_0)^2). \tag{6.1}$$

Indeed, if (6.1) is true, we derive that for n, p large enough,

$$\begin{aligned} P_\theta(t_0 \leq u_\alpha) &= P_\theta(E_\theta t_0 - t_0 \geq E_\theta t_0 - u_\alpha) \leq P_\theta(|E_\theta t_0 - t_0| \geq E_\theta t_0 - u_\alpha) \\ &\leq \frac{\mathrm{Var}_\theta(t_0)}{(E_\theta t_0 - u_\alpha)^2} = o(1) , \end{aligned} \tag{6.2}$$

by Chebychev's inequality. In order to check (6.1), we use the identities

$$E_\theta t_0 = E_X(E_\theta^X t_0), \quad \mathrm{Var}_\theta t_0 = \mathrm{Var}_X(E_\theta^X t_0) + E_X(\mathrm{Var}_\theta^X t_0).$$

Under $P_\theta^X,\ \theta \in \Theta_k(r)$, we have $Y \sim \mathcal{N}_n(v, I_n)$, where

$$v = v(\theta, X) = \sum_{j=1}^p \theta_j X_j, \quad \|v\|^2 = \sum_{j=1}^p \theta_j^2 \|X_j\|^2 + 2 \sum_{1<j<l\le p} \theta_j \theta_l (X_j, X_l).$$

It follows that

$$E_\theta^X(t_0) = (2n)^{-1/2}\|v\|^2, \quad \mathrm{Var}_\theta^X(t_0) = 1 + 2n^{-1}\|v\|^2.$$

Since $E_X(\|X_j\|^2) = n,\ E_X((X_j, X_l)) = 0,\ j \neq l$, we get the first convergence in (6.1):

$$E_\theta t_0 = (2n)^{-1/2} E_X(\|v\|^2) = (n/2)^{1/2} \sum_{j=1}^p \theta_j^2 = (n/2)^{1/2}\|\theta\|^2 \ge (n/2)^{1/2} r^2 \to \infty.$$

Let us turn to the variance term

$$\begin{aligned} E_X(\mathrm{Var}_\theta^X t_0) &= 1 + 2n^{-1} E_X(\|v\|^2) = 1 + 2\|\theta\|^2 = o(E_\theta t_0), \\ \mathrm{Var}_X(E_\theta^X(t_0)) &= (2n)^{-1}\mathrm{Var}_X(\|v\|^2). \end{aligned}$$

By **A2**, the random variables X_{ij} are independent in $(i,j),\ i = 1, ..., n,\ j = 1, .., p$. Consequently, the random variables (X_{j_1}, X_{l_1}) with $\{j_1, l_1\} \neq \{j_2, l_2\}$ are uncorrelated. Moreover, $\|X_{j_1}\|^2$ and (X_j, X_l) are uncorrelated as long as $(j,l) \neq (j_1, j_1)$. We have

$$\mathrm{Var}_X\|X_j\|^2 = \mathrm{Var}_X(X_{ij}^2)n,\ E_X(X_j, X_l)^2 = n,\ j \neq l,$$

where $\mathrm{Var}_X(X_{ij}^2) \le E_X(X_{ij}^4) < \infty$ by **B1**. Then, we get

$$\begin{aligned} n^{-1}\mathrm{Var}_X\|v\|^2 &= n^{-1}\sum_{j=1}^p \theta_j^4 \mathrm{Var}_X\|X_j\|^2 + 4n^{-1} \sum_{1\le j<l\le p} \theta_j^2\theta_l^2 E_X(X_j, X_l)^2 \\ &\le \sup_i [E_X(X_{i1}^4)] \sum_{j=1}^p \theta_j^4 + 4\|\theta\|^4 \le (O(1) + 4)\|\theta\|^4 \\ &= o(n\|\theta\|^4) = o\left((E_\theta t_0)^2\right), \quad \text{as} \quad n\|\theta\|^4 \ge nr^4 \to \infty. \end{aligned}$$

Therefore we get the second relation (6.1).

Note that if $nr^4 \to \infty$, then in the inequality (6.2), we can replace u_α by a sequence $T_{np} \to \infty$ such that $\limsup T_{np} r^{-2} n^{-1/2} < 1$, for instance by $T_{pn} = n^{1/2}r^2/2$. Then, the corresponding test ψ^0 satisfies $\gamma(\psi^0, \Theta_p(r)) \to 0$. Theorem 4.2 follows. $\square$

6.2 Tests based on the statistic t_1

First observe that under H_0, the statistic t_1 is a degenerate U-statistic of the second order, i.e., for $Z_s = (X^{(s)}, Y_s),\ s = 1, 2, 3$ one has $E_{Z_1}K(Z_1, Z_2) = 0$, which yields

$E_0 t_1 = 0$. By Assumption **A1**,

$$\begin{aligned} E_0 t_1^2 &= E_0(K^2(Z_1,Z_2)) = p^{-1}E_0(Y_1^2Y_2^2)\sum_{j=1}^p\sum_{l=1}^p E_X\left(X_{1j}X_{2j}X_{1l}X_{2l}\right) \\ &= p^{-1}\sum_{j=1}^p E_X\left(X_{1j}^2X_{2j}^2\right) = 1. \end{aligned}$$

Set

$$G(Z_1,Z_2) = E_{Z_3}\left(K(Z_1,Z_3)K(Z_2,Z_3)\right),\ G_2 = E_0(G^2(Z_1,Z_2)),\ G_4 = E_0(K^4(Z_1,Z_2)),$$

where E_{Z_3} denotes the expectation over Z_3 under P_0. In order to establish the asymptotic normality of t_1 we only need to check the two following conditions, see [14] Lemma 3.4,

$$G_2 = o(1), \quad G_4 = o(n^2). \tag{6.3}$$

We have by Assumption **A1**,

$$\begin{aligned} G(Z_1,Z_2) &= p^{-1}E_{Z_3}\left(Y_1Y_2Y_3^2\sum_{j=1}^p\sum_{l=1}^p X_{1j}X_{3j}X_{2l}X_{3l}\right) \\ &= p^{-1}Y_1Y_2\sum_{j=1}^p\sum_{l=1}^p X_{1j}X_{2l}E_X(X_{3j}X_{3l}) \\ &= p^{-1}Y_1Y_2\sum_{j=1}^p X_{1j}X_{2j} = p^{-1/2}K(Z_1,Z_2). \end{aligned}$$

Since $E_0(K^2(Z_1,Z_2)) = 1$, we get the first convergence in (6.3). Next by **A2**,

$$\begin{aligned} E_0(K^4(Z_1,Z_2)) &= p^{-2}E_0(Y_1^4Y_2^4)\sum_{j=1}^p\sum_{l=1}^p\sum_{r=1}^p\sum_{s=1}^p E_X(X_{1j}X_{2j}X_{1l}X_{2l}X_{1r}X_{2r}X_{1s}X_{2s}) \\ &= 9p^{-2}\sum_{j=1}^p\sum_{l=1}^p\sum_{r=1}^p\sum_{s=1}^p H_{jlrs}^2, \end{aligned}$$

since $E_0(Y_1^4Y_2^4) = E_0^2(Y_1^4) = 9$, where we set

$$H_{jlrs} \triangleq E_X(X_{1j}X_{1l}X_{1r}X_{1s}) = \begin{cases} E_X(X_1^4), & j=l=r=s, \\ 1, & j=l\neq r=s \text{ or } j=r\neq l=s \text{ or } j=s\neq r=l, \\ 0, & \text{otherwise.} \end{cases}$$

As a consequence, we get

$$E_0(K^4(Z_1,Z_2)) \leq 9p^{-1}b_4^2 + 27\ ,$$

where $b_4 \triangleq \sup_i E(X_{i1}^4)$. By **B1**, the second convergence in (6.3) holds true. Thus, Theorem 4.3 (i) follows.

Let us now evaluate the type II errors under P_θ. Recall that by (1.1),

$$Y_i = \xi_i + v_i, \quad v_i = \sum_{j=1}^{p} \theta_j X_{ij}, \quad \xi_i \sim \mathcal{N}(0,1) \quad iid.$$

Observe that $E_\theta Y_i X_{ij} = \theta_j$ and set

$$K_\theta(Z_1, Z_2) = p^{-1/2} \sum_{j=1}^{p} (Y_1 X_{1j} - \theta_j)(Y_2 X_{2j} - \theta_j).$$

Consider the representation

$$K(Z_1, Z_2) = K_\theta(Z_1, Z_2) + \delta(Z_1) + \delta(Z_2) + h(\theta)$$

where

$$\delta(Z_i) = p^{-1/2} \sum_{j=1}^{p} (Y_i X_{ij} - \theta_j)\theta_j \ , \qquad h(\theta) = p^{-1/2} \sum_{j=1}^{p} \theta_j^2 \ .$$

Observe that the kernel $K_\theta(Z_1, Z_2)$ is symmetric and degenerate under P_θ, i.e.,

$$E_\theta^{Z_1} K_\theta(Z_1, Z_2) = E_\theta^{Z_2} K_\theta(Z_1, Z_2) = 0.$$

The terms $K_\theta(Z_1, Z_2)$, $\delta(Z_1)$, and $\delta(Z_2)$ are centered and uncorrelated under P_θ. As a consequence, we derive that

$$\begin{aligned} E_\theta(K(Z_1, Z_2)) &= p^{-1/2} \|\theta\|^2, & (6.4)\\ \mathrm{Var}_\theta(K(Z_1, Z_2)) &= \mathrm{Var}_\theta(K_\theta(Z_1, Z_2)) + \mathrm{Var}_\theta(\delta(Z_1)) + \mathrm{Var}_\theta(\delta(Z_2)) \ . & (6.5) \end{aligned}$$

Let us compute the variances. Let δ_{ij} be the Kronecker function. Using the representation

$$\begin{aligned} K_\theta(Z_1, Z_2) &= p^{-1/2} \sum_{j=1}^{p} \left(\xi_1 X_{1j} + \sum_{r=1}^{p} \theta_r (X_{1r} X_{1j} - \delta_{rj}) \right) \\ &\qquad \times \left(\xi_2 X_{2j} + \sum_{s=1}^{p} \theta_s (X_{2s} X_{2j} - \delta_{sj}) \right), \end{aligned}$$

we derive that

$$E_\theta^X (K_\theta(Z_1, Z_2)) = p^{-1/2} \sum_{j=1}^{p} \sum_{r=1}^{p} \sum_{s=1}^{p} \theta_r \theta_s (X_{1r} X_{1j} - \delta_{rj})(X_{2s} X_{2j} - \delta_{sj}),$$

Denoting $H_{rsj} = (X_{1r}X_{1j} - \delta_{rj})(X_{2s}X_{2j} - \delta_{sj})$ observe that

$$\mathrm{Var}_X E_\theta^X (K_\theta(Z_1, Z_2)) = p^{-1} \sum_{j=1}^p \sum_{r=1}^p \sum_{s=1}^p \sum_{l=1}^p \sum_{u=1}^p \sum_{v=1}^p \theta_r\theta_s\theta_u\theta_v E_X(H_{rsj}H_{uvl}).$$

Note that

$$E_X(H_{rsj}H_{uvl}) = D_{rujl}D_{svjl},$$

where (we omit the first index $i = 1, 2$ in X_{ij})

$$\begin{aligned} D_{rujl} &= E_X\left((X_rX_j - \delta_{rj})(X_uX_l - \delta_{ul})\right), \\ D_{svjl} &= E_X\left((X_sX_j - \delta_{sj})(X_vX_l - \delta_{vl})\right). \end{aligned}$$

Observe that

$$D_{rujl} = \begin{cases} 1, & r = l \neq u = j \quad \text{or} \quad r = u \neq j = l, \\ b_4 - 1, & r = u = j = l, \\ 0, & \text{otherwise.} \end{cases}$$

We obtain

$$\begin{aligned} &\mathrm{Var}_X E_\theta^X (K_\theta(Z_1, Z_2)) = p^{-1} \sum_{j=1}^p \sum_{r=1}^p \sum_{s=1}^p \sum_{l=1}^p \sum_{u=1}^p \sum_{v=1}^p \theta_r\theta_s\theta_u\theta_v D_{rujl}D_{svjl} \\ &= \frac{(b_4-1)^2}{p} \sum_{r=1}^p \theta_r^4 + \frac{2b_4 - 1}{p} \sum_{1\le r,s\le p,\ r\neq s} \theta_r^2\theta_s^2 + \frac{1}{p} \sum_{1\le j,r,s\le p,\ j\neq r, j\neq s} \theta_r^2\theta_s^2 \\ &= O[\sum_{j=1}^p \theta_j^4] + O[(\sum_{j=1}^p \theta_j^2)^2] = O(\|\theta\|^4). \end{aligned}$$

We now compute $E_X[\mathrm{Var}_\theta^X(K_\theta(Z_1, Z_2))]$.

$$\begin{aligned} \mathrm{Var}_\theta^X(K_\theta(Z_1, Z_2)) &= p^{-1} \sum_{1\le j,l\le p} X_{1j}X_{2j}X_{1l}X_{2l} \\ &+ p^{-1} \sum_{j=1}^p \sum_{l=1}^p \sum_{r=1}^p \sum_{s=1}^p X_{1j}X_{1l}\theta_r\theta_s(X_{2r}X_{2j} - \delta_{jr})(X_{2s}X_{2l} - \delta_{sl}) \\ &+ p^{-1} \sum_{j=1}^p \sum_{l=1}^p \sum_{r=1}^p \sum_{s=1}^p X_{2j}X_{2l}\theta_r\theta_s(X_{1r}X_{1j} - \delta_{jr})(X_{1s}X_{1l} - \delta_{sl}). \end{aligned}$$

Let us take the expectation with respect to X. By Assumption **A2**, we have

$$\begin{aligned} E_X[\mathrm{Var}_\theta^X(K_\theta(Z_1, Z_2))] &= 1 + 2p^{-1} \sum_{j=1}^p \sum_{r=1}^p \sum_{s=1}^p \theta_r\theta_s E_X[(X_{2r}X_{2j} - \delta_{jr})(X_{2s}X_{2j} - \delta_{sj})] \\ &\le 1 + 2\sum_{r=1}^p b_4\theta_r^2 = 1 + O(\|\theta\|^2) = O(1 + \|\theta\|^4) \end{aligned}$$

Since

$$\mathrm{Var}_\theta(K_\theta(Z_1,Z_2)) = E_X \mathrm{Var}_\theta^X(K_\theta(Z_1,Z_2)) + \mathrm{Var}_X E_\theta^X(K_\theta(Z_1,Z_2)),$$

we get

$$\mathrm{Var}_\theta(K_\theta(Z_1,Z_2)) = O(1+\|\theta\|^4). \tag{6.6}$$

Similarly for $i=1,2$, we compute the variance of $\delta(Z_i)$.

$$\delta(Z_i) = p^{-1/2}\sum_{j=1}^p \theta_j\left(\xi_i X_{ij} + \sum_{l=1}^p \theta_l (X_{ij}X_{il} - \delta_{jl})\right),$$

and we have (we omit the index $i=1,2$)

$$\begin{aligned}
E_\theta^X(\delta(Z)) &= p^{-1/2}\sum_{j=1}^p\sum_{l=1}^p \theta_j\theta_l(X_jX_l-\delta_{jl}),\\
\mathrm{Var}_\theta^X(\delta(Z)) &= p^{-1}\sum_{j=1}^p\sum_{l=1}^p \theta_j\theta_l X_jX_l, \quad E_X\mathrm{Var}_\theta^X(\delta(Z)) = p^{-1}\|\theta\|^2,\\
\mathrm{Var}_X E_\theta^X(\delta(Z)) &= p^{-1}\sum_{j=1}^p\sum_{l=1}^p\sum_{r=1}^p\sum_{s=1}^p \theta_j\theta_l\theta_r\theta_s D_{jrls}\ ,
\end{aligned}$$

where D_{jrls} was previously defined and upper bounded. This yields

$$\mathrm{Var}_X E_\theta^X(\delta(Z)) = p^{-1}\left((b_4-1)\sum_{j=1}^p \theta_j^4 + 2\|\theta\|^4\right) = O(\|\theta\|^4/p). \tag{6.7}$$

Combining (6.4), (6.5), (6.6), and (6.7) we obtain, for $r^2np^{-1/2} \to \infty$ and $p=o(n^2)$,

$$\begin{aligned}
E_\theta(t_1) &= \sqrt{N}E_\theta(K(Z_1,Z_2)) = \sqrt{N}h(\theta) \sim n(2p)^{-1/2}\|\theta\|^2 \geq \frac{n}{\sqrt{2p}}r^2 \to \infty,\\
\mathrm{Var}_\theta(t_1) &= \mathrm{Var}_\theta(K_\theta(Z_1,Z_2)) + \frac{n^3}{N}\mathrm{Var}_\theta(\delta(Z_1)) = O(1+\|\theta\|^4) + O(n\|\theta\|^4/p)\\
&= o\left((E_\theta(t_1))^2\right).
\end{aligned}$$

Applying Chebyshev's inequality as in the proof of Theorem 4.2 allows to conclude. □

6.3 Higher Criticism Tests

6.3.1 Type I errors

The variables $X_1,...,X_p,Y$ are independent under P_0 and $(X_j,a)/\|a\| \sim \mathcal{N}(0,1)$ for any $a \in \mathbb{R}^p,\ a \neq 0$ under **A3**. Thus we have

$$\begin{aligned}
P_0(y_1<t_1,...,y_p<t_p) &= E_Y(P_0^Y((X_1,Y)/\|Y\|<t_1,...,(X_p,Y)/\|Y\|<t_p)\\
&= E_Y(\Phi(t_1)....\Phi(t_p)) = \Phi(t_1)....\Phi(t_p) = P_0(y_1<t_1)...P_0(y_p<t_p).
\end{aligned}$$

It follows that $y_j = (X_j, Y)/\|Y\| \sim \mathcal{N}(0,1)$ and $y_1, \ldots, y_p$ are i.i.d. under P_0. As a consequence, the random variables q_i are independent uniformly distributed on $(0,1)$ under P_0. We denote by $F_p(t)$ the empirical distribution of $(q_i)_{1\leq i\leq p}$:

$$F_p(t) = \frac{1}{p}\sum_{i=1}^{p} 1\!\!1_{q_i \leq t} \ .$$

Then, the normalized uniform empirical process is defined by

$$W_p(t) = \sqrt{p}\frac{F_p(t) - t}{\sqrt{t(1-t)}} \ .$$

Arguing as in Donoho and Jin [8], we observe that $t_{HC} = \sup_{t\leq 1/2} W_p(t)$. It is stated in [26], Chapter 16 that

$$\frac{\sup_{0\leq t\leq 1/2} W_p(t)}{\sqrt{2\log\log p}} \rightarrow_P 1 \ , \quad p\rightarrow\infty \ .$$

This proves the result. □

6.3.2 Type II errors

We define $H_{np} = (1+a)\sqrt{2\log\log p}$. Consider some $\beta \in (1/2, 1)$ and assume that $k\log(p)/n \rightarrow 0$. It is sufficient to prove that for any $\delta_0 > 0$ arbitrarily small the radius

$$r_{np} = (\varphi(\beta) + \delta_0)\sqrt{k\log(p)/n} \tag{6.8}$$

satisfies

$$\beta(\psi^{HC}, \Theta_k(r_{np})) \rightarrow 0 \ . \tag{6.9}$$

For any $\theta \in \Theta_k$, we set $\|\theta\|_\infty \triangleq \sup_i |\theta_i|$. In order to prove the convergence (6.9), we consider a partition of $\Theta_k(r_{np})$:

$$\begin{aligned}
\tilde{\Theta}_k^{(1)}(r_{np}) &\triangleq \Theta_k(r_{np}) \cap \left\{\theta \in \Theta_k, \|\theta\|^2 \geq \frac{4k\log(p)}{n}\right\} \\
\tilde{\Theta}_k^{(2)}(r_{np}) &\triangleq \Theta_k(r_{np}) \cap [\tilde{\Theta}_k^{(1)}(r_{np})]^c \cap \left\{\theta \in \Theta_k \ , \|\theta\|_\infty^2 \geq \frac{4\log(p)}{n}\right\} \\
\tilde{\Theta}_k^{(3)}(r_{np}) &\triangleq \Theta_k(r_{np}) \cap [\tilde{\Theta}_k^{(1)}(r_{np})]^c \cap [\tilde{\Theta}_k^{(2)}(r_{np})]^c \ .
\end{aligned}$$

The sets $\tilde{\Theta}_k^{(1)}(r_{np})$ and $\tilde{\Theta}_k^{(2)}(r_{np})$ contain the parameters θ whose l_2 or l_∞ norms are large, while the set $\tilde{\Theta}_k^{(3)}(r_{np})$ contains the remaining parameters.

Proposition 6.1 *Consider the set of parameters $\tilde{\Theta}_k^{(4)}$ defined by*

$$\tilde{\Theta}_k^{(4)} \triangleq \left\{\theta \in \Theta_k, \quad \frac{\|\theta\|_\infty^2}{1+\|\theta\|^2} \geq \frac{3\log(p)}{n}\right\} \ .$$

Let us introduce the statistic $t_{\max}$ and the corresponding test $\psi^{\max}$ defined by

$$t_{\max} \triangleq (pq_{(1)})^{-1/2} - (pq_{(1)})^{1/2} \leq t_{HC} \ , \qquad \psi^{\max} \triangleq \mathbb{1}_{t_{\max} > H_{np}} \ .$$

We have $\beta(\psi^{\max}, \tilde{\Theta}_k^{(4)}) \to 0$.

It follows that $\beta(\psi^{HC}, \tilde{\Theta}_k^{(4)}) \to 0$. Observe that

$$\tilde{\Theta}_k^{(1)}(r_{np}) \subset \left\{ \theta \in \Theta_k \ , \quad \frac{\|\theta\|^2}{1+\|\theta\|^2} \geq \frac{4k\log(p)/n}{1+4k\log(p)/n} \right\} \ .$$

Since $k\|\theta\|_\infty^2 \geq \|\theta\|^2$ and since $k\log(p)/n$ converges to 0, it follows that $\tilde{\Theta}_k^{(1)}(r_{np}) \subset \tilde{\Theta}_k^{(4)}$ for n large enough. Thus, we get $\beta(\psi^{HC}, \tilde{\Theta}_k^{(1)}(r_{np})) \to 0$.

Let us turn to $\tilde{\Theta}_k^{(2)}(r_{np})$. For any $\theta \in \tilde{\Theta}_k^{(2)}(r_{np})$, we have

$$\frac{\|\theta\|_\infty^2}{1+\|\theta\|^2} \geq \frac{4\log(p)/n}{1+4k\log(p)/n} \ .$$

This quantity is larger than $3\log(p)/n$ for n large enough. We get $\beta(\psi^{HC}, \tilde{\Theta}_k^{(2)}(r_{np})) \to 0$.

Proposition 6.2 *Let us set $T_p = \sqrt{\log(p)}$ and $u > 0$ such that*

$$u = \begin{cases} 2\varphi(\beta), & \beta \in (1/2, 3/4] \ , \\ \sqrt{2}, & \beta \in (3/4, 1) \ . \end{cases} \tag{6.10}$$

We consider the statistic $L(u)$ and the corresponding test ψ^L defined by

$$L(u) \triangleq \sum_{j=1}^{p} \frac{\mathbb{1}_{|y_j| > uT_p} - 2\Phi(-uT_p)}{\sqrt{2p\Phi(-uT_p)}} \ , \qquad \psi^L = \mathbb{1}_{L(u) \geq H_{np}} \ .$$

Then, $\beta(\psi^L, \tilde{\Theta}_k^{(3)}(r_{np})) \to 0$. Moreover, we have $L(u) \leq t_{HC}$, for p large enough.

It follows from Proposition 6.2 that $\beta(\psi^{HC}, \tilde{\Theta}_k^{(3)}(r_{np})) \to 0$ converges to 0, which completes the proof. $\square$

6.3.3 Proof of Proposition 6.1

It follows directly from the definition (4.7) that $t_{\max} \leq t_{HC}$. Consider the test $\psi'^{\max}$ defined by

$$\psi'^{\max} = \mathbb{1}_{\|y\|_\infty \geq \sqrt{2.5\log(p)}} \tag{6.11}$$

If $\psi'^{\max} = 1$, it follows that $q_{(1)} \leq 2\Phi(-\sqrt{2.5\log(p)}) \leq 2p^{-5/4}$. Hence, we have $t_{\max} \geq p^{1/8}/\sqrt{2} - \sqrt{2}p^{-1/8}$. For p large enough, this implies that $\psi^{\max} = 1$.

Consequently, we only have to prove that $\beta(\psi'^{\max}, \tilde{\Theta}_k^{(4)}) \to 0$.

Consider $\theta \in \tilde{\Theta}_k^{(4)}$. By symmetry, we may assume that $\|\theta\|_\infty = |\theta_1|$. We use the following decomposition

$$\|Y\|y_1 = \theta_1\|X_1\|^2 + (Y - \theta_1 X_1, X_1) \ .$$

The random variables $\|Y\|^2/(1+\|\theta\|^2)$ and $\|X_1\|^2$ have a χ^2 distribution with n degrees of freedom. Since $Y - \theta_1 X_1$ is independent of X_1, the random variable $(Y - \theta_1 X_1, X_1/\|X_1\|)$ is normal with mean 0 and variance $1 + \sum_{i\neq 1}\theta_i^2$.

With probability larger than $1 - O(n^{-1} \vee \log^{-1}(p))$, we obtain

$$\begin{aligned}
\|Y\|^2/n &\leq (1+\|\theta\|^2)[1 + o(n^{-1/4})] \\
(1 - o(n^{-1/4})) \leq \|X_1\|^2/n &\leq (1 + o(n^{-1/4})) \\
|(Y - \theta_1 X_1, X_1)|/\|X_1\| &\leq (1 + \sum_{i\neq 1}\theta_i^2)^{1/2}\sqrt{2\log(\log(p))} \ .
\end{aligned}$$

Thus, we get

$$|y_1| \geq \frac{\sqrt{n}|\theta_1|}{(1+\|\theta\|^2)^{1/2}}[1 - o(n^{-1/4})] - O(\sqrt{\log\log(p)}) \ ,$$

with probability larger than $1 - O(n^{-1} \vee \log^{-1}(p))$. Since $\theta \in \tilde{\Theta}_k^{(4)}$, we have $n|\theta_1|^2/(1+\|\theta\|^2) \geq 3\log(p)$ and the test $\psi'_{\max}$ rejects with probability going to one. It follows that $\beta(\psi'^{\max}, \tilde{\Theta}_k^{(4)}) \to 0$. $\square$

6.3.4 Proof of Proposition 6.2

Connection between t_{HC} and $L(u)$. Set $\hat{s}_u \triangleq \sum_{i=1}^p \mathbb{1}_{|y_j|>uT_p}$. Observe that $q_{(\hat{s}_u)} \leq P(|\mathcal{N}(0,1)| > uT_p) \leq 1/2$ for p large enough. If follows that

$$L(u) = \frac{\sqrt{p}[\hat{s}_u/p - 2\Phi(-uT_p)]}{\sqrt{2\Phi(-uT_p)}} \leq \frac{\sqrt{p}[\hat{s}_u/p - q_{(\hat{s}_u)}]}{\sqrt{q_{(\hat{s}_u)}}} \leq t_{HC} \ .$$

Power of ψ^L. Under P_θ, $\|Y\|^2/(1+\|\theta\|^2)$ has a χ^2 distribution with n degrees of freedom. For any $\theta \in \tilde{\Theta}_k^{(3)}(r_{np})$, we have $\|\theta\|^2 \leq 4k\log(p)/n = o(1)$. As a consequence, we have $|\|Y\|^2 - n| \leq 4k\log(p) + 4\sqrt{n\log(n)} = o(n)$ with probability larger than $1 - O(1/n)$ uniformly over all $\theta \in \tilde{\Theta}_k^{(3)}(r_{np})$. Consider the event $\mathcal{Z}_{np,1} = \{|\|Y\|^2 - n| \leq H_n\}$, where $H_n = 4k\log(p) + 4\sqrt{n\log(n)} = o(n)$. It is sufficient to prove that

$$\sup_{\theta\in\tilde{\Theta}_k^{(3)}(r_{np})} P_\theta(\mathcal{Z}_{np,1} \cap \{L(u) \leq H_{np}\}) \to 0. \tag{6.12}$$

Consider $\theta \in \tilde{\Theta}_k^{(3)}(r_{np})$. We can assume that $\theta_{k+1} = ... = \theta_p = 0$. Then $Y = \sum_{j=1}^k \theta_j X_j + \xi$ does not depend on $X_{k+1}, ..., X_p$. Arguing as for the type I error, we derive that $y_{k+1},...,y_p$ are independent standard Gaussian variables and do not depend on $(y_1,...,y_k)$. We can write $L(u) = L_1(u) + L_2(u)$, where

$$\begin{aligned} L_1(u) &= \frac{\sum_{j=1}^k \left(1\!\!{\rm I}_{\{|y_j|>uT_p\}} - 2\Phi(-uT_p)\right)}{\sqrt{2p\Phi(-uT_p)}}, \\ L_2(u) &= \frac{\sum_{j=k+1}^p \left(1\!\!{\rm I}_{\{|y_j|>uT_p\}} - 2\Phi(-uT_p)\right)}{\sqrt{2p\Phi(-uT_p)}}. \end{aligned}$$

We find

$$E_\theta(L_2(u)) = 0, \quad \mathrm{Var}_\theta(L_2(u)) = \frac{2p\Phi(-uT_p)(1-2\Phi(-uT_p))}{2p\Phi(-uT_p)} \le 1 \ ,$$

which yields,

$$P_\theta(|L_2(u)| > H_{np}) \to 0 \ . \tag{6.13}$$

In order to study the term $L_1(u)$, we will find a statistic $\tilde{L}_1(u)$ such that $P_\theta[\tilde{L}_1(u) < L_1(u)] = 1 + o(1)$ uniformly over $\Theta_k^{(3)}(r_{np})$. For such a $\tilde{L}_1(u)$, we will have

$$P_\theta[L(u) \le H_{np}] \le P_\theta[L_1(u) \le 2H_{np}] + o(1) \le P_\theta[\tilde{L}_1(u) \le 2H_{np}] + o(1). \tag{6.14}$$

Construction of $\tilde{L}_1(u)$. Observe that under P_θ,

$$\begin{aligned} y_j &= (\hat{y}_j\|\xi\| + n\theta_j + \Delta_j)/\|Y\| \ , \\ \Delta_j &= \sum_{l\neq j}^k \theta_l(X_j, X_l) + \left(\|X_j\|^2 - n\right)\theta_j, \quad j = 1,...,k, \end{aligned}$$

where

$$\hat{y}_j = (X_j, \xi)/\|\xi\|.$$

We only need to consider $Z \in \mathcal{Z}_{np,2} = \{|\|\xi\|^2 - n| < n^{2/3}\}$ since $P_\theta(\mathcal{Z}_{np,2}) \to 1$. Set $\mathcal{Z}_{np,3} = \mathcal{Z}_{np,1} \cap \mathcal{Z}_{np,2}$. Thus, for $\delta = \delta_{np} \to 0$ one has

$$\begin{aligned} \{|y_j| > uT_p\} \cap \mathcal{Z}_{np,3} &\supset \{|n^{-1/2}\hat{y}_j\|\xi\| + n^{-1/2}\Delta_j + n^{1/2}\theta_j)| > uT_p(1+\delta)\} \cap \mathcal{Z}_{np,3} \\ &\supset \{sgn(\theta_j)\hat{y}_j(1-\delta) > uT_p(1+\delta) - n^{1/2}|\theta_j| + |\tilde{S}_j|)\} \cap \mathcal{Z}_{np,3} \ , \end{aligned} \tag{6.15}$$

where $\tilde{S}_j = n^{-1/2}\Delta_j$.

Lemma 6.1 *For any $T > 0$ going to infinity and such that $T = o(\sqrt{n})$, we have*

$$\log(P_X(|\tilde{S}_j| > T\|\theta\|)) \le -\frac{1}{4}T^2(1 + o(1)) \ ,$$

uniformly over $\in \tilde{\Theta}_k^{(3)}(r_{np})$.

Taking $T = \sqrt{4\log(p)}$, we obtain

$$P_X(|\tilde{S}_j| > T\|\theta\|) = o(p^{-1}).$$

We recall that $\|\theta\|^2 \le 4k\log(p)/n = o(1)$ since $\theta \in \tilde{\Theta}_k^{(3)}(r_{np})$. Hence, we get

$$P_X\left[\max_{1\le j\le k} |\tilde{S}_j| > o(\sqrt{\log(p)})\right] = o(1), \quad \text{uniformly over } \theta \in \tilde{\Theta}_k^{(3)}(r_{np}).$$

Combining this bound with (6.15), we obtain that there exists an event $\mathcal{Z}_{np,4}$ of probability tending to one and a sequence $\delta = \delta_{np} \to 0$ such that

$$\{|y_j| > uT_p\} \cap \mathcal{Z}_{np,4} \supset \{sgn(\theta_j)\hat{y}_j > uT_p(1+\delta) - (1-\delta)n^{1/2}|\theta_j|\} \cap \mathcal{Z}_{np,4}\ . \quad (6.16)$$

Observe that the random variables $\hat{y}_j$ are independent standard normal.

Setting $\tilde{u} = u(1+\delta)$, $\tilde{\rho}_j = (1-\delta)n^{1/2}|\theta_j|$ we define

$$\tilde{L}_1(u) = \frac{\left(\sum_{j=1}^k \mathbb{1}_{\hat{y}_j > \tilde{u}T_p - \tilde{\rho}_j} - 2\Phi(-uT_p)\right)}{\sqrt{2p\Phi(-uT_p)}}.$$

By (6.16), $\tilde{L}_1(u)$ satisfies $P_\theta[\tilde{L}_1(u) \le L_1(u)] = 1 - o(1)$ uniformly over $\tilde{\Theta}_k^{(3)}(r_{np})$. In view of (6.14), in order to complete the proof it suffices to show that

$$P_\theta[\tilde{L}_1(u) \le 2H_{np}] = o(1) \quad \text{uniformly over } \tilde{\Theta}_k^{(3)}(r_{np}). \quad (6.17)$$

Control of $P_\theta[\tilde{L}_1(u) \le 2H_{np}]$. In order to evaluate this probability, recall that $\hat{y}_j \sim \mathcal{N}(0,1)$ i.i.d. under P_θ. Thus,

$$\begin{aligned} E_\theta(\tilde{L}_1(u)) &= \frac{\left(\sum_{j=1}^k \Phi(-\tilde{u}T_p + \tilde{\rho}_j) - 2\Phi(-uT_p)\right)}{\sqrt{2p\Phi(-uT_p)}}\ , \\ \mathrm{Var}_\theta(\tilde{L}_1(u)) &\le \frac{\sum_{j=1}^k \Phi(-\tilde{u}T_p + \tilde{\rho}_j)}{2p\Phi(-uT_p)}\ . \end{aligned}$$

By Chebyshev's inequality, we get

$$\begin{aligned} P_\theta(\tilde{L}_1(u) \le 2H_{np}) &= P_\theta(E_\theta(\tilde{L}_1(u)) - \tilde{L}_1(u) \ge E_\theta(\tilde{L}_1(u)) - 2H_{np}) \\ &\le \frac{\mathrm{Var}_\theta(\tilde{L}_1(u))}{(E_\theta(\tilde{L}_1(u)) - 2H_{np})^2}. \end{aligned}$$

Lemma 6.2 *There exists $\eta > 0$ such that, for n, p large enough,*

$$\inf_{\theta\in\tilde{\Theta}_k^{(3)}(r_{np})} \frac{\sum_{j=1}^k \Phi(-\tilde{u}T_p + \tilde{\rho}_j)}{\sqrt{2p\Phi(-uT_p)}} \sim \inf_{\theta\in\tilde{\Theta}_k^{(3)}(r_{np})} E_\theta(\tilde{L}_1(u)) > p^\eta. \quad (6.18)$$

In the sequel, we denote by A_p a log-sequence, i.e., a sequence such that $A_p = (\log(p))^{c_p}$, $|c_p| = O(1)$ as $p \to \infty$. Since $u \in [0, \sqrt{2}]$, we have $p\Phi(-uT_p) \ge A_p$. Combining this bound with Lemma 6.2 yields

$$\mathrm{Var}_\theta(\tilde{L}_1(u)) = O\left(A_p E_\theta(\tilde{L}_1(u))\right).$$

Since $H_{np} = o(p^\eta)$, this implies (6.17) and then (6.12). $\square$

6.3.5 Proof of Lemma 6.1

Let us bound the deviations of $\tilde{S}_j$ by computing the exponential moments of Δ_j. For any h such that $h^2\|\theta\|^2 \leq 1/4$, we have

$$\begin{aligned}
E_X(\exp(h\Delta_j)) &= E_{X_j} E_X^{X_j}(\exp(h\Delta_j)) \\
&= E_{X_j}\left(\exp\left(h\theta_j(\|X_j\|^2 - n\right) E_X^{X_j} \exp\left(h\sum_{l\neq j}\theta_l(X_j, X_l)\right)\right) \\
&= E_{X_j}\left(\exp\left(h\theta_j(\|X_j\|^2 - n) + \frac{h^2}{2}\|X_j\|^2 \sum_{l\neq j}\theta_l^2\right)\right) \\
&= \exp\left(-nh\theta_j - \frac{n}{2}\log\left(1 - 2h\theta_j - h^2\sum_{l\neq j}\theta_l^2\right)\right),
\end{aligned}$$

as $2h\theta_j + h^2\sum_{l\neq j}\theta_l^2 < 1$. Taking h such that $h^2 k\log(p) = o(n)$ enforces $h^2\|\theta\|^2 = o(1)$. Using the Taylor expansion of the logarithm

$$-hx - \frac{1}{2}\log(1 - 2hx - h^2y^2) = \frac{1}{2}h^2(2x^2 + y^2)(1 + o(1)),\ h^2(2x^2 + y^2) = o(1),$$

we get

$$\begin{aligned}
E_X(\exp(h\Delta_j)) &= E_X(\exp(\sqrt{n}h\tilde{S}_j)) = \exp\left(\frac{n}{2}h^2(2\theta_j^2 + \sum_{l\neq j}\theta_l^2)(1 + o(1))\right) \\
&\leq \exp\left[nh^2\|\theta\|^2(1 + o(1))\right] \qquad (6.19)
\end{aligned}$$

as $h^2\|\theta\|^2 = o(1)$. Take some $T > 0$. Applying a standard technique based on Markov's inequality yields

$$\begin{aligned}
P_X(\tilde{S}_j > T\|\theta\|) &\leq E_X\left[\exp\left(\frac{T\tilde{S}_j}{2\|\theta\|} - \frac{T^2}{2}\right)\right], \\
P_X(-\tilde{S}_j > T\|\theta\|) &\leq E_X\left[\exp\left(-\frac{T\tilde{S}_j}{2\|\theta\|} + \frac{T^2}{2}\right)\right].
\end{aligned}$$

We get from (6.19) that

$$\log(P_X(|\tilde{S}_j| > T\|\theta\|)) \leq -\frac{1}{4}T^2(1 + o(1)) \quad \text{if} \quad T^2 = o(n) \text{ and } T \to \infty.$$

□

6.3.6 Proof of Lemma 6.2

Recall that we consider $r_{np} = (\varphi(\beta) + \delta_0)\sqrt{k\log(p)/n}$ with arbitrarily small $\delta_0 > 0$ (see (6.8)). Recalling that $T_p = \sqrt{\log(p)}$, we apply the results of Section 7.5 for $\delta = \delta_{np} > 0,\ \delta_{np} = o(1)$, and

$$T = \tilde{u}T_p, \quad \tilde{u} = (1 + \delta)u, \quad v = (1 - \delta)(\varphi(\beta) + \delta_0) < \tilde{u}, \quad t_0 = vT_p, \quad R = 2T_p,$$

since for $t_j = \tilde{\rho}_j$ one has

$$\sum_{j=1}^{k} t_j^2 = (1-\delta)^2 n \sum_{j=1}^{k} \theta_j^2 \geq (1-\delta)^2 n r_{np}^2 = k t_0^2.$$

By the choice of u, and v, the relations (7.4) hold true for p large enough (see Remark 7.1). Applying Lemmas 7.4 and 7.5, we get

$$\inf_{\theta \in \tilde{\Theta}_k(r_{np})} \sum_{j=1}^{k} \Phi(-\tilde{u}T_p + \tilde{\rho}_j) = k\Phi(-\tilde{u}T_p + t_0).$$

We recall that A_p denotes any log-sequence. Since $\Phi(-tT_p) = A_p p^{-t^2/2}$ for $t > 0$, we have

$$\begin{aligned}
\inf_{\theta \in \tilde{\Theta}_k(r_{np})} E_\theta(\tilde{L}_1(u)) &= \frac{k\left(\Phi(-\tilde{u}T_p + t_0) - 2\Phi(-uT_p)\right)}{\sqrt{2p\Phi(-uT_p)}} \sim \frac{k\Phi(-\tilde{u}T_p + t_0)}{\sqrt{2p\Phi(-uT_p)}} \\
&= \frac{k\Phi(-(\tilde{u}-v)T_p)}{\sqrt{2p\Phi(-uT_p)}} = A_p p^{1/2-\beta-(\tilde{u}-v)_+^2/2+u^2/4}.
\end{aligned}$$

In order to obtain (6.18), we have to check that there exists $\eta > 0$ such that, for n, p large enough,

$$G \triangleq \frac{1}{2} - \beta - \frac{(\tilde{u}-v)_+^2}{2} + \frac{u^2}{4} \geq \eta.$$

Let $\beta \in (1/2, 3/4]$. Recalling that $\varphi^2(\beta) = 2\beta - 1 > 0$ and (6.10) we see, that for $\delta = \delta_{np} = o(1)$ and $\delta_0 \in (0, \varphi(\beta))$, one can find $\eta = \eta(\beta, \delta_0) > 0$ such that

$$\begin{aligned}
G &= -\frac{\varphi^2(\beta)}{2} - \frac{(\varphi(\beta)-\delta_0)^2}{2} + \varphi^2(\beta) + o(1) \\
&= \varphi(\beta)\delta_0 - \frac{\delta_0^2}{2} + o(1) \geq \eta + o(1).
\end{aligned}$$

Let us now consider $\beta \in (3/4, 1]$. Recalling that $\varphi(\beta) = \sqrt{2}(1-\sqrt{1-\beta})$ and (6.10), we see that for $\delta = \delta_{np} = 0(1)$ and $\delta_0 \in (0, \sqrt{2-2\beta})$, one can find $\eta = \eta(\beta, \delta_0) > 0$ such that

$$\begin{aligned}
G &= \frac{1}{2} - \beta - \frac{\left(\sqrt{2} - \sqrt{2}\left(1-\sqrt{1-\beta}\right) - \delta_0\right)^2}{2} + \frac{1}{2} + o(1) \\
&= 1 - \beta - \left(\sqrt{1-\beta} - \delta_0/\sqrt{2}\right)^2 + o(1) \\
&= \sqrt{2-2\beta}\,\delta_0 - \frac{\delta_0^2}{2} + o(1) \geq \eta + o(1).
\end{aligned}$$

The relation (6.18) follows. $\square$

6.4 Proof of Proposition 4.6

Under H_0, the distributions of the variables $(y_i)_{i=1,\ldots,p}$ do not depend on σ^2. As a consequence, $E_{0,\sigma}(\psi^{HC}) = E_{0,1}$. This last quantity has been shown to converge to 0 in Theorem 4.4. Hence, we get $\alpha^{un}(\psi^{HC}) = o(1)$.

Let us turn to the type II error probability. We consider the model $Y_i = \sum_{j=1}^{p} \theta_j X_{ij} + \xi_i$ where $\mathrm{Var}(\xi_i) = \sigma^2$. Dividing this equation by σ, we obtain the model:

$$Y_i' = \sum_{j=1}^{p} (\theta_j/\sigma) X_{ij} + \xi_i' \ ,$$

where $\mathrm{Var}(\xi_i') = 1$. The statistic t_{HC} is exactly the same for the data $Z = (Y, X)$ and $Z' = (Y', X)$. Consequently, we obtain $E_{\theta\sigma,\sigma}(1 - \psi^{HC}) = E_{\theta,1}(1 - \psi^{HC})$. It remains to use the bound on $E_{\theta,1}(1 - \psi^{HC})$ from Theorem 4.4. $\square$

7 Appendix: Technical results

7.1 Thresholds

Take the thresholds $T = T_j$ satisfying

$$T_j = \frac{a_j}{2} + \frac{\log(h^{-1})}{a_j}.$$

Define $a_j = x_j\sqrt{\log(p)}$, $\tau_j = T_j/\sqrt{\log(p)}$, and $h = p^{-\beta}$. Then, we have $\tau_j = x_j/2 + \beta/x_j$.

If for some $\delta_0 > 0$, $x_j + \delta_0 < \varphi_2(\beta) \stackrel{\Delta}{=} \sqrt{2}(1 - \sqrt{1-\beta}) \leq \varphi(\beta)$, then there exists $\delta_1 > 0$ such that $\tau_j > \sqrt{2} + \delta_1$. For such a x_j, we derive that

$$pT_j^r \Phi(-T_j) = o(1), \quad \forall\ r > 0\ . \tag{7.1}$$

In particular, if $x_j = o(1)$, then $\tau_j \to \infty$ and (7.1) holds.

For any $\delta > 0$, we have

$$\Phi(-T_j) \asymp h\Phi(-T_j + a_j) \quad \text{for} \quad \tau_j > x_j + \delta\ . \tag{7.2}$$

This holds if $x_j < \varphi_2(\beta) \leq \sqrt{2}$.

7.2 Norms $\|X_j\|$ and scalar products (X_j, X_l)

Clearly,

$$E(\|X_j\|^2) = n, \quad E(X_j, X_l) = 0\ , \quad \mathrm{Var}(X_j, X_l) = n.$$

By Assumption **B1**, there exists $D > 0$ such that $\sup_{j\neq l} \mathrm{Var}(X_j, X_l) \leq nD$ and $\sup_j \mathrm{Var}(\|X_j\|^2) \leq nD$.

Lemma 7.1 *Let U_j be a random variable distributed as X_{ij}.*

(1) Assume that there exists $h_0 > 0$ such that $\sup_{1\le j\le l\le p} E(e^{hU_jU_l}) < \infty$ for any $|h| < h_0$. Then, for any sequence $t = t_n$ such that $t = o(\sqrt{n})$ and $t\sqrt{n} \to \infty$,

$$P(|\|X_j\|^2 - n| > t\sqrt{n}) \;\; \le \;\; \exp[-t^2/(2D)(1 + o(1))],$$

and

$$P(|(X_j, X_l)|| > t\sqrt{n}) \;\; \le \;\; \exp[-t^2/(2D)(1 + o(1))],$$

(2) Assume that $E(|X|^m) < \infty$, for some $m > 2$. Then there exists $C_m < \infty$ such that

$$P(|\|X_j\|^2 - n| > t\sqrt{n}) \le C_m t^{-m/2}, \quad P(|(X_j, X_l)| > t\sqrt{n}) \le C_m t^{-m}.$$

Proof follows from the standard arguments based on the moment inequalities and exponential inequalities. If $EZ = 0,\ \mathrm{Var}(Z) = 1,\ E(e^{h_0 Z}) < \infty$, then $\log(Ee^{hZ}) = h^2/2(1 + o(1))$ as $h \to 0$. Hence, we take $h = t/\sqrt{n} = o(1)$ for the study of the exponential moments of $S_n = \sum_{i=1}^n Z_i$.

Corollary 7.1

(1) Let $\log(p) = o(n)$ and the assumptions Lemma 7.1 (1) hold true. Then, for any $B > 2$, one has

$$P_X(\max_{1\le j\le p} |\|X_j\|^2 - n| > \sqrt{BDn\log(p)}\,) = o(1),$$

$$P_X(\max_{1\le j<l\le p} |(X_j, X_l)| > \sqrt{2BDn\log(p)}\,) = o(1).$$

(2) Let $p = o(n^{m/4})$ and the assumptions Lemma 7.1 (2) hold true. Then, for any sequence v_n going to infinity, one has

$$P_X(\max_{1\le j\le p} |\|X_j\|^2 - n| > \sqrt{n} p^{2/m} v_n) = o(1),$$

$$P_X(\max_{1\le j<l\le p} |(X_j, X_l)| > \sqrt{n} p^{2/m} v_n) = o(1).$$

(3) Under assumptions (1) or (2) uniformly in $1 \le j < l \le p$ in P_X-probability, one has $a_j \sim a = b\sqrt{n}, \quad x_j \sim x$, i.e., for any $\delta > 0$,

$$P_X(\max_{1\le j\le p} |(a_j/b\sqrt{n}) - 1| > \delta) \to 0, \quad P_X(\max_{1\le j\le p} |(x_j/x) - 1| > \delta) \to 0.$$

7.3 Expansion of $\Phi(t)$

Let $\Phi(t)$ be the standard Gaussian cdf and $\phi(t)$ be the standard Gaussian pdf.

Lemma 7.2 *Let $\delta \to 0,\ t\delta = O(1)$. Then*

$$\Phi(t + \delta) = \Phi(t) + \delta\phi(t) + O\left(\delta^2(|t| + 1)\phi(t)\right).$$

Proof follows from the Taylor expansion and the properties of $\phi(t)$. □

Observe that for any $b \in \mathbb{R}$ there exists $C = C(b) > 0$ such that $(|t|+1)\phi(-t) \le C(b)\Phi(-t)$ as $t \le b$. It follows from Lemma 7.2 that as $\delta \to 0,\ t\delta = O(1),\ t \le B$ for some $B \in \mathbb{R}$, then

$$\Phi(-t + \delta) = \Phi(-t)(1 + O(\delta^2)) + \delta\phi(t).$$

7.4 Tails of correlated vectors

Lemma 7.3 *Let (X,Y) be the Gaussian random two-dimensional vector,*

$$EX = EY = 0,\ \mathrm{Var}(X) = \mathrm{Var}(Y) = 1,\ \mathrm{Cov}(X,Y) = r.$$

Let $t_1 \asymp t_2 \to \infty,\ rt_1 = o(1)$. Then

$$P(X > t_1, Y > t_2) = \Phi(-t_1)\Phi(-t_2)\left(1 + O(r^2)\right) + r\phi(t_1)\phi(t_2).$$

Proof. Observe that the conditional distribution $\mathcal{L}(Y|X = x)$ is Gaussian $\mathcal{N}(m(x), \sigma^2(x))$ with $m(x) = rx,\ \sigma^2(x) = 1 - r^2$. Therefore

$$P(X > t_1, Y > t_2) = \int_{t_1}^{\infty} P(Y > t_2|X = x)d\Phi(x) = \int_{t_1}^{\infty} \Phi\left(\frac{-t_2 + rx}{\sqrt{1-r^2}}\right) d\Phi(x).$$

Setting $h = |r|^{-1}$, observe that

$$\int_h^{\infty} \Phi\left(\frac{-t_2 + rx}{\sqrt{1-r^2}}\right) d\Phi(x) \leq \Phi(-h) = o\left(r^2\Phi(-t_1)\Phi(-t_2)\right).$$

It is sufficient to study the integral over the interval $\Delta = [t_1, h]$. For $x \in \Delta$, we have

$$\frac{-t_2 + rx}{\sqrt{1-r^2}} = -t_2 + \delta(x), \quad \delta(x) = rx + O(r^2 t_2 + |r^3 x|)) = O(1).$$

Applying Lemma 7.2, we have

$$\begin{aligned}
&\int_{\Delta} \Phi\left(\frac{-t_2 + rx}{\sqrt{1-r^2}}\right) d\Phi(x) \\
&= \Phi(-t_2)\left(\Phi(-t_1) - \Phi(-h)\right)\left(1 + O(r^2)\right) + r\phi(t_2)\int_{\Delta} x d\Phi(x) \\
&= \Phi(-t_1)\Phi(-t_2)\left(1 + O(r^2)\right) + r\phi(t_1)\phi(t_2),
\end{aligned}$$

since $\int_{\Delta} x d\Phi(x) = \phi(t_1) - \phi(h) = \phi(t_1) + o(r^2\Phi(-t_1))$. $\square$

7.5 A minimization problem

Let $f(t)$ be a function defined on the interval $t \in [0, R]$. Consider the minimization problem

$$F_k(t_0) = \inf \sum_{j=1}^{k} f(t_j) \quad \text{subject to} \quad \sum_{j=1}^{k} t_j^2 \geq k t_0^2, \quad t_j \in [0, R]. \tag{7.3}$$

Lemma 7.4 *Assume that there exists $\lambda > 0$ such that*

$$\inf_{t \in [0,R]} (f(t) - \lambda t^2) = f(t_0) - \lambda t_0^2.$$

Then $F_k(t_0) = k f(t_0)$.

Proof. We have, for any $(t_1, ..., t_k)$ such that $t_j \in [0, R]$, $\sum_{j=1}^k t_j^2 \geq kt_0^2$,

$$\begin{aligned}\sum_{j=1}^k f(t_j) &\geq \sum_{j=1}^k f(t_j) - \lambda \left(\sum_{j=1}^k t_j^2 - kt_0^2 \right) = \sum_{j=1}^k \left(f(t_j) - \lambda t_j^2 \right) + k\lambda t_0^2 \\ &\geq k \left(f(t_0) - \lambda t_0^2 \right) + k\lambda t_0^2 = kf(t_0). \quad \square\end{aligned}$$

We apply Lemma 7.4 to the function $f(x) = \Phi(-T + x)$. Let $\phi(x) = \Phi'(x)$ stand for the standard Gaussian pdf.

Lemma 7.5 *Let $f(t) = \Phi(-T + t)$. Suppose*

$$t_0 > 0, \quad T > t_0 + \frac{2}{t_0}, \quad T < R \leq \left(\frac{t_0}{\phi(-T + t_0)} \right)^{1/2}. \tag{7.4}$$

Take $\lambda = \phi(-T + t_0)/2t_0$. Then the assumptions of Lemma 7.4 are fulfilled, i.e.,

$$\inf_{0 \leq t \leq R} (f(t) - \lambda t^2) = f(t_0) - \lambda t_0^2.$$

Proof. Denote $g(t) = \Phi(-T + t) - \lambda t^2$. By the choice of λ we have $g'(t_0) = 0$. Let us consider the second derivative,

$$g''(t) = (T - t)\phi(-T + t) - 2\lambda = (T - t)\phi(-T + t) - \phi(-T + t_0)/t_0.$$

Observe that the function $-x\phi(x)$ is positive for $x < 0$, increases for $x \in (-\infty, -1)$ and decreases for $x \in (-1, 1)$; $\lim_{x \to -\infty} \phi(x) = 0 = 0\phi(0)$,

$$g''(t_0) = (T - t_0 - t_0^{-1})\phi(-T + t_0) > 0.$$

Consequently, there exist two points t_1, t_2 such that $t_1 < t_0 < t_2 < T$,

$$g''(t_1) = g''(t_2) = 0, \quad g''(t) < 0 \quad \text{as} \quad t < t_1 \quad \text{and} \quad t > t_2.$$

The function $g(t)$ is therefore convex on $[t_1, t_2]$, concave on $(-\infty, t_1]$ and on $[t_2, \infty)$, and t_0 is the point of a local minimum of $g(t)$. By the concavity, this yields that the global minimum of $g(t)$ at $t \in [0, R]$ is achieved either at $t = t_0$ or at the ends of the interval $[0, R]$. Therefore we only need to show that $g(0) > g(t_0)$ and $g(R) > g(t_0)$.

In order to verify the first inequality, observe that $g(0) > 0$. Recalling the well known inequality:

$$\Phi(-y) < \frac{1}{y}\phi(-y), \ \forall y > 0 \ ,$$

we get

$$g(t_0) = \Phi(-T + t_0) - t_0\phi(-T + t_0)/2 < \phi(-T + t_0) \left(\frac{1}{T - t_0)} - \frac{t_0}{2} \right) < 0 \ ,$$

because $T > t_0 + 2t_0^{-1}$.

The second inequality follows from the relation

$$g(R) = \Phi(-T + R) - \frac{R^2\phi(-T + t_0)}{2t_0} > \frac{t_0 - R^2\phi(-T + t_0)}{2t_0} > 0,$$

in view of the assumption on R. $\square$

Remark 7.1 Observe that if $0 < v < u < b$ and

$$T = uT_p, \quad t_0 = vT_p, \quad R = bT_p \ ,$$

where T_p is large enough, then assumptions (7.4) hold.

References

[1] Aldous, D. J. (1985). *Exchangeability and Related Topics.* Lecture Notes in Mathematics **1117**, Springer-Verlag.

[2] Arias-Castro, E., Candès E. J., and Plan Y. (2010) Global Testing and Sparse Alternatives: ANOVA, Multiple Comparisons and the Higher Criticism. arXiv:1007.1434.

[3] Bickel, P. J., Ritov, Y. and Tsybakov, A. B. (2009). Simultaneous analysis of Lasso and Dantzig selector. *Ann. Statist.*, **37**, 1705–1732.

[4] Cai, T., Jin, J. and Low, M. (2007) Estimation and confidence sets for sparse normal mixtures. *Ann. Statist.* **35**, 2421–2449.

[5] Candès, E. and Tao, T. (2007). The Dantzig selector: statistical estimation when p is much larger than n. *Ann. Statist.*, **35** 2313–2351.

[6] Davidson, K. R. and Szarek, S. J. (2001). Local operator theory, random matrices and Banach spaces. In *Handbook of the geometry of Banach spaces, Vol. I*, 317–366.

[7] Donoho, D.L. (2006) Compressed Sensing. *IEEE Transactions on Information Theory*, **52** 1289–1306.

[8] Donoho, D. and Jin, J. (2004) Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32**, 962-994.

[9] Donoho, D. and Jin, J. (2008) Higher criticism thresholding: Optimal feature selection when useful features are rare and weak. *Proc. Nat. Acad. Sci.* **105**, 14790-14795.

[10] Donoho, D. and Jin, J. (2009) Feature selection by higher criticism thresholding achieves optimal phase diagram. *Royal Society Philosophical Transactions A* **367**, 4449-4470.

[11] Hall, P. and Jin, J. (2010) Innovated higher criticism for detecting sparse signals in correlated noise. *Annals of Statistics* **38**, 1686-1732.

[12] Haupt, J., Castro, R. and Nowak, R. (2008) Adaptive discovery of sparse signals in noise. In Proc. 42th Asilomar Conference on Signal, Systems and Computers, Pacific Grove, California, October 2008.

[13] Haupt, J., Castro, R. and Nowak, R. (2008) Distilled sensing: Adaptive sampling for sparse detection and estimation. arXiv:1001.5311.

[14] Ingster, Yu.I. (1994) Minimax hypothesis testing on a distribution density for ellipsoids in l_p. *Theory Probab. Appl.*, **39**, 530–553.

[15] Ingster, Yu.I. (1997) Some problems of hypothesis testing leading to infinitely divisible distributions. *Math. Methods of Stat.* **6**, 47-69.

[16] Ingster, Yu. I. and Suslina, I. A. (2001*a*) Adaptive detection of a signal of growing dimension. I. *Math. Methods of Stat.* **10**, 395-421.

[17] Ingster, Yu. I. and Suslina, I. A. (2001*b*) Adaptive detection of a signal of growing dimension. II. *Math. Methods of Stat.* **11**, 37-68.

[18] Ingster, Yu. I. and Suslina, I. A. (2002*a*) *Nonparametric goodness-of-fit testing under gaussian models*. Springer Lectures Notes in Statistics. vol. 169.

[19] Ingster, Yu. I. and Suslina, I .A. (2002*b*) On a detection of a signal of known shape in multichannel system. *Zapiski Nauchn. Sem. POMI* **294**, 88-112 (Transl. *J. Math. Sci.* **127**, 1723-1736).

[20] Ingster, Yu. I., Pouet, C. and Tsybakov, A. B. (2009) Sparse classification boundaries. arXiv:0903.4807.

[21] Ingster, Yu. I., Pouet, C. and Tsybakov, A. B. (2009) Classification of sparse high-dimensional vectors. *Royal Society Philosophical Transactions A* **367**, 4427–4448.

[22] Jager, L. and Wellner, J. A. (2007) Goodness-of-fit tests via phi-divergences. *Ann. Statist.* **35**, 2018-2053.

[23] Jin, J. (2003) Detection boundary for sparse mixtures. Ph.D. thesis, Department of Statistics, Stanford University.

[24] Jin, J. (2004) Detecting a target in very noisy data from multiple looks. In *A festschrift for Herman Rubin*. IMS Lecture Notes Monogr. Ser., no. 45, pp. 255–286.

[25] Petrov, V.V. (1995) *Limit theorems of probability theory*. Oxford: Oxford University Press.

[26] Shorack, G. and Wellner, J. (1986) *Empirical processes with applications to statistics*. Wiley Series in Probability and Mathematical Statistics.

[27] Verzelen, N. and Villers, F. (2010) Goodness-of-fit tests for high-dimensional Gaussian linear models. *Annals of Statistics*, **38**, 704-752.

[28] Verzelen, N. (2010) Minimax risks for sparse regressions: Ultra-high-dimensional phenomenons. arXiv:1008.0526.

[29] Wainwright, M.J.(2009a) Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (Lasso). *IEEE Trans. Information Theory*, **55**, 2183–2202.

[30] Wainwright, M.J.(2009b) Information-Theoretic Limits on Sparsity Recovery in the High-Dimensional and Noisy Setting. *IEEE Trans. Information Theory*, **55**, 5728-5741.